

For Reference

NOT TO BE TAKEN FROM THIS ROOM

Ex libris
UNIVERSITATIS
ALBERTAENSIS





Digitized by the Internet Archive
in 2020 with funding from
University of Alberta Libraries

<https://archive.org/details/Singh1971>

THE UNIVERSITY OF ALBERTA

BAYESIAN STATISTICAL METHODS
IN MODEL DISCRIMINATION AND MODEL BUILDING

by



Harbhajan Singh

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE
OF MASTER OF SCIENCE

DEPARTMENT OF CHEMICAL AND PETROLEUM ENGINEERING

EDMONTON, ALBERTA

SPRING, 1971

THE UNIVERSITY OF CHICAGO

DEPARTMENT OF CHEMISTRY

PHYSICAL CHEMISTRY

PHYSICAL CHEMISTRY (3)

PHYSICAL CHEMISTRY (3) is a course in physical chemistry. It covers the topics of thermodynamics, statistical mechanics, and quantum mechanics. The course is designed for students who have completed the introductory course in physical chemistry. The course is taught by Professor [Name] and is held in the Department of Chemistry, University of Chicago.

PHYSICAL CHEMISTRY (3) is a course in physical chemistry. It covers the topics of thermodynamics, statistical mechanics, and quantum mechanics. The course is designed for students who have completed the introductory course in physical chemistry. The course is taught by Professor [Name] and is held in the Department of Chemistry, University of Chicago.

PHYSICAL CHEMISTRY (3)

PHYSICAL CHEMISTRY (3)

THE UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies for acceptance a thesis entitled BAYESIAN STATISTICAL METHODS IN MODEL DISCRIMINATION AND MODEL BUILDING submitted by Harbhajan Singh in partial fulfilment of the requirements for the degree of Master of Science.



ABSTRACT

The modeling of a physical phenomenon is basic both to science and engineering. While little scientific thought has gone towards the problem of model formulation, substantial effort has been devoted to the process of selecting the best model from among the competing models to describe a physical phenomenon. It is intended here to introduce some improvements and standardize the statistical model discrimination procedure.

Present expressions for the expected likelihood of data involve working with the entire set of data. New expressions are developed that utilize only the sufficient statistics of the data thereby affecting substantial savings in computation and avoiding large round off errors. Justification of the conventional least squares follows as a special case of the Bayesian method from these new developments.

Frequently, it is necessary to perform further experiments to discriminate between several rival models. Hitherto, no exact criterion has been available for the sequential design of experiments so as to obtain maximum information in picking the right model. An expected entropy change criterion in line with Shannon's concept of entropy as a function of information about the state of the system, is presented here.

Finally, certain practices of Bayesian procedure are critically examined from the conceptual viewpoint. Examples of model building in Petroleum Engineering and kinetics are presented in the last chapter.

ACKNOWLEDGEMENTS

It is a pleasure to express my sincere appreciation to Professor D. Quon whose guidance and helpful criticism made this work possible.

I also wish to express my thanks to Dr. R. G. Bentsen for his advice in the field of petroleum reservoir parameter estimation; and to the National Research Council of Canada for the financial aid.

TABLE OF CONTENTS

1.	INTRODUCTION	
1.1	Mathematical Models for Real World	1
1.2	Modeling Process	1
1.3	Review of the Previous Work in Model Discrimination	6
2.	EXPECTED LIKELIHOODS OF DATA	21
2.1	Development of the Expected Likelihood of the Data from the Sufficient Statistics	21
3.	SEQUENTIAL DESIGN OF EXPERIMENTS FOR MODEL DISCRIMINATION	30
3.1	Development of Expected Entropy Change Criterion for Model Discrimination	30
3.2	Examples	37
4.	STANDARDIZATION OF THE BAYESIAN MODEL DISCRIMINATION PROCEDURE	48
4.1	Examination of the Bayesian Procedure	48
4.2	Rationalized Bayesian Model Discrimination Procedure	63
5.	APPLICATION EXAMPLES	
5.1	Petroleum Reservoir Estimation from Material Balance	65
5.2	Petroleum Reservoir Estimation - Field Examples	77
5.3	Parameter Estimation from a Set of Non-linear Differential Equations	85
6.	DISCUSSION	92
	BIBLIOGRAPHY	96
	APPENDIX	A-1

1. INTRODUCTION

1.1 Mathematical Models for the Real World

A mathematical model is an abstract formulation aimed to simulate the mechanism of a real world phenomenon. A true mathematical model accounting completely for the physical phenomenon is a hypothetical concept and is seldom realized in non-trivial situations. Nevertheless, it is a useful tool for the scientists and engineers in representing the important characteristics of the actual mechanism. An analyst is often concerned with a situation where a dependent variable y , is related to a set of independent variables $\underline{x} = (x_1, x_2, \dots, x_n)$. For example, a stock broker might be studying the dependency of a company's share value y , on the total investment of the company x_1 , the capital investment rate per year x_2 , the company's total annual sales x_3 , the bank lending interest rate x_4 , etc. Theoretically, there exists a relation of the type

$$y = \phi(\underline{x}, \underline{a}) \quad (1.1)$$

between the dependent variable y and the set of independent variables $\underline{x}$, where $\underline{a} = (a_1, a_2, \dots, a_m)$ is the set of m parameters which may be the physical constants of the system under study. Usually, a simpler relation that closely approximates the outstanding features of (1.1) is satisfactory.

A mathematical model that describes the mechanism reasonably well might also be useful to the scientists in compressing a large

amount of experimental data into a single structure. In addition, the model may provide new insight to the phenomenon, suggest further experiments and improvements in the model, the purpose here being to establish the true nature of the process. For an engineer, on the other hand, the quantitative description is usually required for predictive purposes. The mathematical model is a tool, whereby the engineer can employ analytical techniques in the solution of his problems. With the help of the model, he can isolate the important aspects of the problem and investigate alternate designs.

However, all models do not have to be mechanistic models. Depending upon the purpose of the investigation, an empirical relation might be entirely satisfactory. It would be a waste of resources to try to obtain a mechanistic model where the response to the set of independent variables is quite smooth and the only purpose is to have a working design in the region of interest (RI), particularly when extensive data is available in the region. A polynomial type empirical model of the form

$$y = \sum_{i=1}^m a_i \cdot \phi_i(x_i) \quad (1.2)$$

will generally be satisfactory and all that is required is to furnish a suitable routine to generate successive functions to achieve the desired accuracy. Response surface studies deal primarily with the empirical formulation of the models [see for example, Davies (1953), Box (1954, 1960, 1964), Hunter (1958, 1959a, 1959b)].

In other situations where the primary interest is not merely to predict the response over a limited region but rather to elucidate the mechanism or to obtain meaningful results in regions relatively unknown, the concern of the investigator is to obtain a mechanistic representation. This kind of model can provide useful information where the optimization of design leans heavily on the knowledge of what actually is taking place. The mechanistic models themselves are seldom exact, and the results in far off regions may not be reliable. However, an empirical model lends absolutely no assurance to extrapolation.

The subject matter of this thesis is concerned mainly with the technique of obtaining the best possible mechanistic model. The objectives are to devise methods for: (a) obtaining the posterior model probabilities, (b) design of experiments for discriminating among rival models, (c) examination of the consistency of the existing Bayesian model discrimination techniques.

1.2 Modeling Process

The utility of the mathematical models to the scientists and engineers make the modeling process itself an important operation. Smallwood (1968) illustrates the conventional iterative process of model building as shown in Figure 1(a). The first step is to formulate the requirements of the model. As we have seen, if the interest of the engineer is merely to have an approximate response over a finite region of interest, a simple empirical model will do. Alternatively, resort to a complex mechanistic model might be necessary.

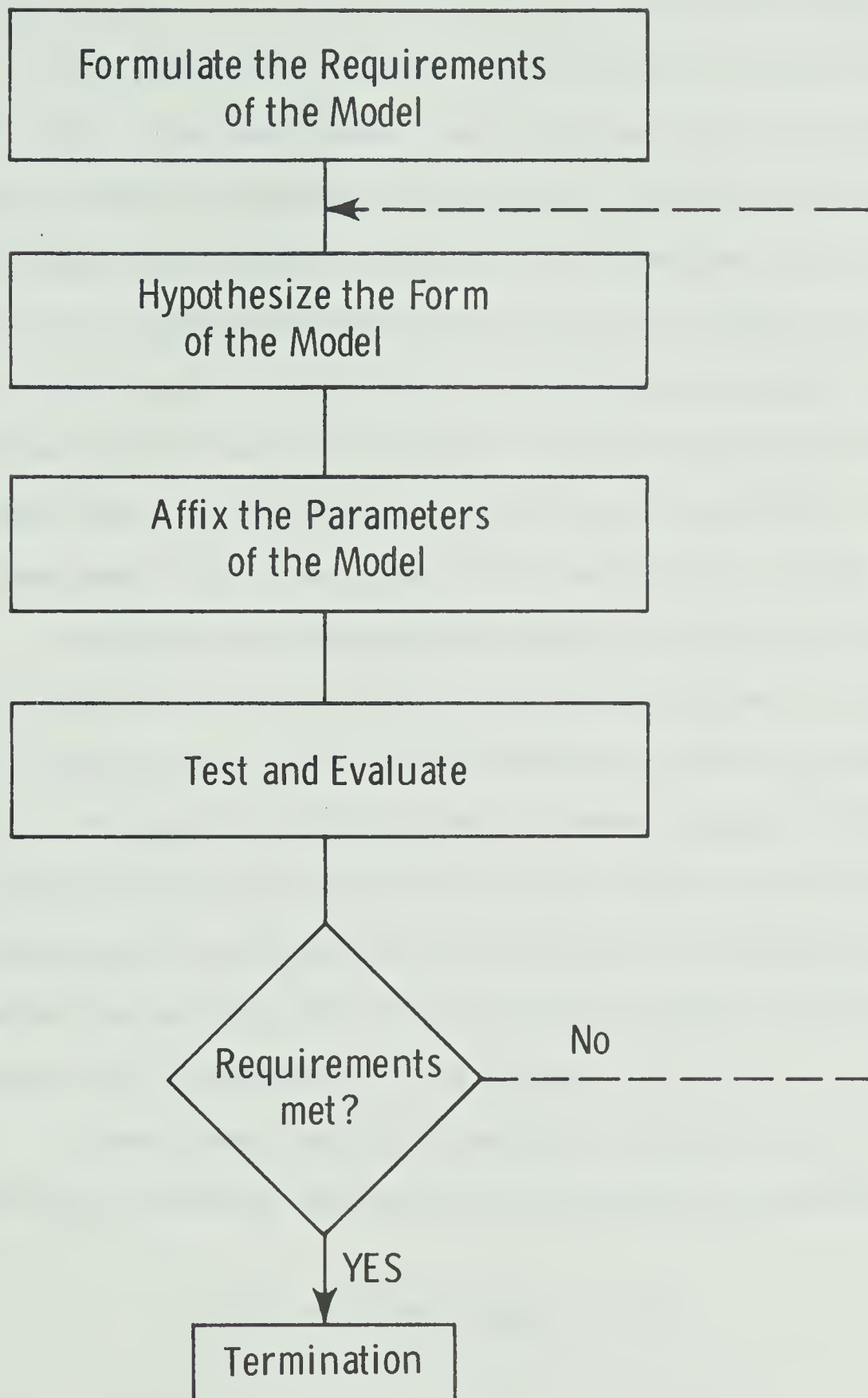


FIGURE 1.1 (a) MODELING PROCESS

A trade-off between the computational complexity and the accuracy of the representation is essential during the course of this step.

The second step involves hypothesizing the form of the model. This is the most crucial step since the form of the model is directly related to the physical mechanism. A close consultation between the statistician and the scientist or engineer is required. The difficulty might be overcome by training the engineer to be his own statistician. Once a form is ascribed to the model, the next step is to estimate its parameters against the available data. The model, then is tested against the original requirements. An alternate model form can be tried if the requirements are not met.

Deciding on the form of the model as in step two above, is to recognize the salient features of the process and is called process identification. The subsequent step of ascribing certain values to the parameters of the model is termed parameter estimation. The process of selecting the best parameter values can be viewed more concretely by developing a cost function of the parameter values. The objective function, then is to minimize the cost of using a particular set of parameters in the design.

Given a model form ϕ_j is decided, the total cost of using a particular estimate $\hat{\underline{a}}$ for the set of parameters can be expressed as

$$Z(\hat{\underline{a}}_j) = \int_{-\infty}^{\infty} f(\underline{a}_j) \cdot Z(\hat{\underline{a}}_j, \underline{a}_j) \cdot d\underline{a}_j \quad (1.3)$$

where $f(\underline{a}_j)$ is the joint distribution of the parameters $\underline{a}_j$ given the model form ϕ_j is true and $Z(\hat{\underline{a}}_j, \underline{a})$ is the cost of using $\hat{\underline{a}}$ as an

estimate of the parameters when the actual value is $\underline{a}$.

The cost of using a particular set of parameters may be interpreted as the savings that could be realized in terms of eliminating overdesign had the parameter values been known for certain. This, then represents the additional amount that could be spent for having the complete information about the system. Stated differently, it is the price one would be prepared to pay for eliminating uncertainty associated with the situation through additional data acquisition. The economic balance between the cost of the experiments and the savings realized in the design due to reduced uncertainty is the key in deciding the extent of experimentation.

However, there frequently arise situations, where many model forms describe the phenomenon about equally well. In such cases, it might be desirable to consider all such model forms simultaneously, with each model form having a set of unknown parameters. This is specifying a model-space called a metamodel [Smallwood (1968)] which contains all eligible models, each with all possible values of the parameters. Process identification and parameter estimation are now handled as a joint problem. The termination of the model discrimination process, in this case, is decided by the cost function, that is the expectation of the cost function associated with all the m individual model forms.

$$Z = \frac{\sum_{j=1}^m p'_j \int_{-\infty}^{\infty} f(\underline{a}_j) \cdot L(\underline{y}|\underline{a}_j) \cdot Z(\hat{\underline{a}}_j, \underline{a}) \cdot d\underline{a}_j}{\sum_{j=1}^m p'_j \int_{-\infty}^{\infty} f(\underline{a}_j) \cdot L(\underline{y}|\underline{a}_j) \cdot d\underline{a}_j} \quad (1.4)$$

where $L(\underline{y}|\underline{a}_j)$ is the likelihood of observing the data given the j th model form is true with parameters $\underline{a}_j$ and p'_j is the prior probability of the j th model being true. Figure 1.1(b) illustrates the modeling process with the metamodel approach.

The concept of a metamodel has been used in Chapter 3 to develop a simple and useful expression for the discriminating function.

1.3 Review of the Previous Work in Model Discriminations

The model discrimination procedure regarded as the process of selecting the best model from the various eligible models, is well-established. Considerable work has been done in this field. After the preliminary screening of the data to eliminate the poor models, model discrimination may be categorized in the following three areas: (a) parameter estimation, (b) computing the model discrimination index--likelihoods of the data, and (c) sequential design of experiments. These areas will be reviewed briefly to serve as an introduction to the text of this thesis.

1.3.1 Review of the Previous Work in Model Discriminations

A comprehensive treatment of parameter estimation of linear and non-linear models is given by Lindley (1965), Graybill (1961) and Raiffa and Schlaifer (1961). Lee (1968) and Donnelly and Quon (1970) have described parameter estimation procedures when the models are in the form of differential equations.

The Bayesian approach to parameter estimation has direct advantages over the least squares method in that the prior knowledge about the parameters can be effectively utilized in the Bayesian method.

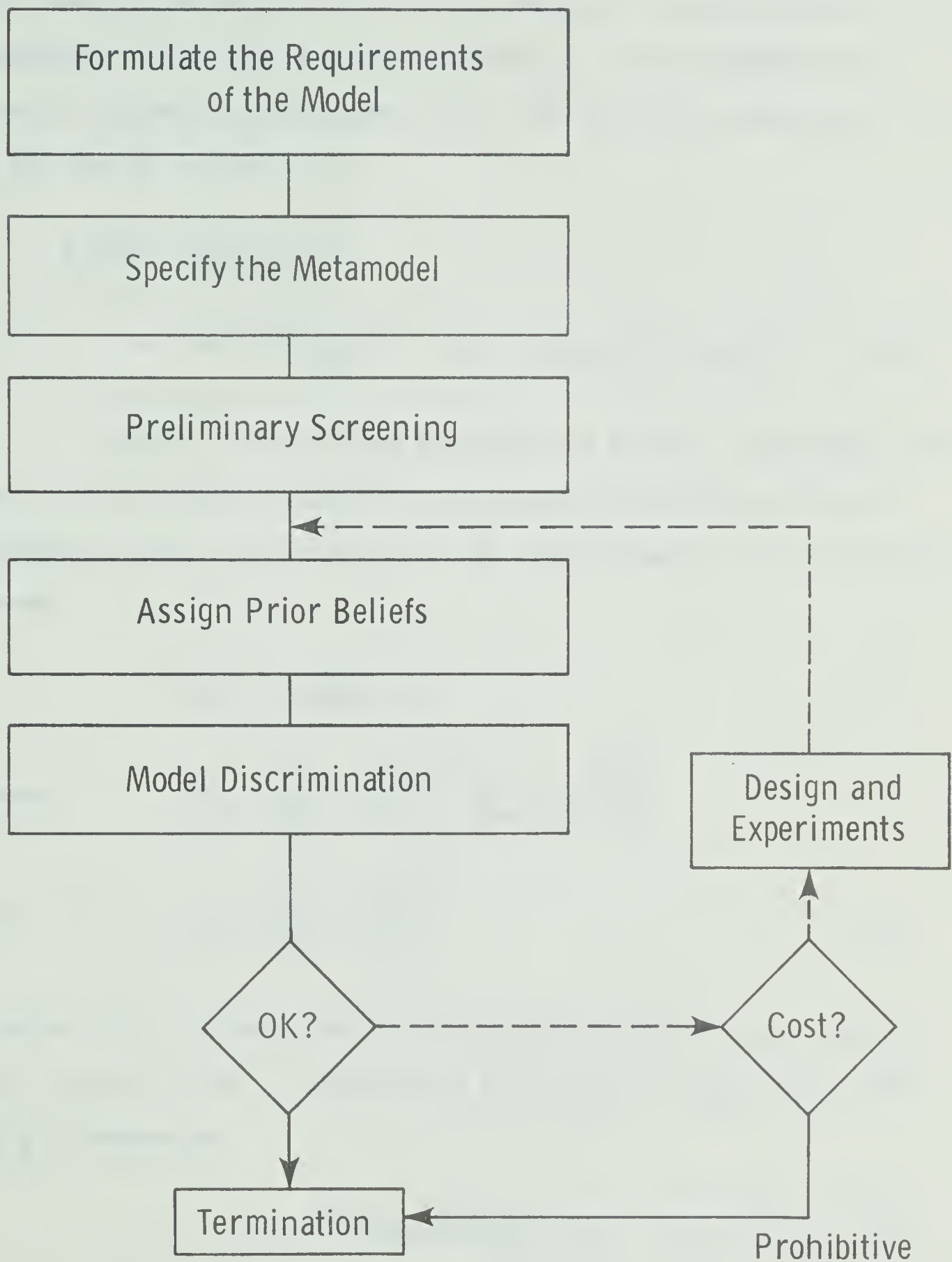


FIGURE 1.1 (b) METAMODELING PROCESS

For example, the beliefs about the parameters of a model may be expressed in the form of prior estimates $\underline{\mu}_0$, of the parameters and a prior variance-covariance matrix $\underline{C}_0$. The prior distribution of $\underline{a}$ may then be written as

$$\begin{aligned} f'(\underline{a}) &= f_N(\underline{a} | \underline{\mu}_0, \underline{C}_0) \\ &= (2\pi)^{-1/2} \cdot |\underline{C}_0|^{-1/2} \cdot \exp\left\{-\frac{1}{2} [\underline{a} - \underline{\mu}_0]^t \underline{C}_0^{-1} [\underline{a} - \underline{\mu}_0]\right\} \end{aligned} \quad (1.5)$$

Now if the errors are assumed to be normally distributed with mean zero and known variance σ^2 , the posterior distribution of the parameters after n observations may be approximated by the multivariate normal

$$f''(\underline{a}) = f_N(\underline{a} | \underline{\mu}_n, \underline{C}_n)$$

where

$$\underline{\mu}_n = \left[\underline{C}_0^{-1} + \frac{\underline{X}^t \underline{X}}{\sigma^2} \right]^{-1} \cdot \left[\underline{C}_0^{-1} \underline{\mu}_0 + \frac{\underline{X}^t \underline{Y}}{\sigma^2} \right]$$

and

$$\underline{C}_n = \left[\underline{C}_0^{-1} + \frac{\underline{X}^t \underline{X}}{\sigma^2} \right]^{-1} \quad (1.6)$$

$\underline{X}$ being the $n \times m$ design matrix of the partial derivatives of $\phi(\underline{x}, \underline{a})$ with respect to the ' m ' parameters $\underline{a} = (a_1, a_2, \dots, a_m)$. An element of $\underline{X}$ is denoted by

$$x_{ij} = \left[\partial \phi(\underline{x}_i, \underline{a}) / \partial a_j \right]_{\underline{a} = \underline{a}_0} \quad \begin{array}{l} i = 1, 2, \dots, n \\ j = 1, 2, \dots, m \end{array} \quad (1.7)$$

where $\underline{a}_0$ is a guessed value of $\underline{a}$ and $\underline{x}_i$ is the experimental condition for the i th observation.

In addition to incorporating the prior knowledge about the parameters, the Bayesian method gives direct reliability estimates of the parameters in terms of the posterior variance-covariance matrix.

For the case when error variance is not known, a joint estimation of the variance and the parameters can be obtained by Bayesian analysis [see Raiffa and Schlaifer (1961), Quon (1970)]. However some sort of estimate is obtained for the error variance and in subsequent calculations, this estimate is treated as if it were a known quantity. The relative utility of the various estimates of the error variance is examined in Chapter 4, from the view point of model discrimination.

Sometimes the investigator may want to determine the parameters of a model with the maximum possible precision. Box and Lucas (1959) have proposed a criterion to select n data point settings that will give the most precise parameters. The criterion maximizes the determinant of $\underline{X}^t \underline{X}$. The maximization of the determinant of $\underline{X}^t \underline{X}$ corresponds to the minimization of m -dimensional volume of the confidence region in the parameter space. Directly, it leads to a minimal contribution in the variance covariance matrix of the model, thereby increasing the precision of the parameters. A chemical engineering problem has been studied using the minimum volume design principle by Kittrel, Hunter and Watson (1966).

A Bayesian procedure for estimating common parameters from several responses was developed by Box and Draper (1965). The method is shown to give more precise estimates of the parameters than the conventional least squares. Draper and Hunter (1968) developed a

sequential design criterion to obtain the most precise estimate of parameters for more than one response.

1.3.2 Model Likelihoods

Least sum of error squares is the discriminating criterion for the common least squares technique. A model form that gives the lowest least sum of error squares is taken to be the true model and the remaining models are discarded. Justification of favouring a particular model by this primitive method is given in Chapter 2 from a Bayesian statistical analysis.

It seems more logical to compute the probability (called the likelihood) of the occurrence of the data given that a particular model is true, and then draw inferences from these likelihoods. The likelihood of observing the given data conditional upon j th model being true with parameters $\underline{a}_j$ is given by [Raiffa and Schlaifer (1961), Lindley (1965)]

$$L(\underline{y}|\underline{a}_j, \phi_j) = (2\pi\sigma_j^2)^{-n/2} \cdot \exp\left\{-\frac{\hat{S}_j^2}{2\sigma_j^2}\right\} \cdot \exp\left\{-\frac{1}{2} [\underline{a}_j - \hat{\underline{a}}_j]^t \frac{\underline{X}_j^t \underline{X}_j}{\sigma_j^2} [\underline{a}_j - \hat{\underline{a}}_j]\right\} \quad (1.8)$$

where n is the number of observations, σ_j^2 is the error variance for the j th model and includes both the measurement error and the modeling error, $\hat{S}_j^2$ is the minimum sum of error squares for the j th model and $\hat{\underline{a}}$ is the corresponding parameter vector that results in minimum sum

of errors squared. $\underline{X}$ is the design matrix described in 1.3.1.

In the maximum likelihood procedure, the maximum value of the likelihoods from (1.8) are computed for each of the models and is taken as a measure of the relative likelihood of the model. This approach too has the obvious drawback in that it does not take into account any asymmetry in the likelihood function. Besides, the prior information about the individual model form is generally not utilized in the maximum likelihood procedure.

The Bayesian procedure is to compute the expected likelihoods of the data for a particular model form, over the entire parameter space $\underline{a}_j$.

$$L(\underline{y}|\phi_j) = \int_{-\infty}^{\infty} L(\underline{y}|\phi_j, \underline{a}_j) \cdot f(\underline{a}_j) \cdot d\underline{a}_j \quad (1.9)$$

The symbol $L(\underline{y}|\phi_j)$ is used to denote the expected value of the random variable $L(\underline{y}|\phi_j, \underline{a}_j)$. In obtaining the expected likelihood of the model, one may use the prior or posterior distribution of $\underline{a}_j$. There are arguments for and against the use of the either distribution. A discussion on the subject is provided in Chapter 4.

As an example, using the posterior distribution of the parameters and a uniform prior for the parameters, the expected likelihood is [see Plackett (1960)]

$$\begin{aligned}
L(\underline{y}|\phi_j) &= f_N^{(n)}(\underline{y}|\underline{X}_j\hat{\underline{a}}, \sigma_j^2[\underline{X}_j\underline{A}_j^{-1}\underline{X}_j^t + \underline{I}]) \\
&= (2\pi)^{-n/2} \cdot |\underline{X}_j\underline{C}_j\underline{X}_j^t + \sigma_j^2\underline{I}| \\
&\quad \exp\left\{-\frac{1}{2} [\underline{y}-\underline{X}_j\hat{\underline{a}}_j]^t [\underline{X}_j\underline{C}_j\underline{X}_j^t + \sigma_j^2\underline{I}]^{-1} [\underline{y}-\underline{X}_j\hat{\underline{a}}_j]\right\} \quad (1.10)
\end{aligned}$$

where

$$\underline{A}_j = \underline{X}_j^t \underline{X}_j \text{ and } \underline{C}_j = \frac{1}{\sigma_j^2} \underline{A}_j^{-1} \quad (1.11)$$

$\underline{X}_j$ being the $n \times m$ design matrix for the j th model, $\underline{I}$ is an $n \times n$ identity matrix and $|\underline{A}|$ in general is written for the determinant of $\underline{A}$.

The exponent of (1.10) may be computed by noting that
[Scheffe (1959)]

$$\begin{aligned}
&[\underline{y}-\underline{X}_j\hat{\underline{a}}_j]^t [\underline{X}_j\underline{C}_j\underline{X}_j^t + \sigma_j^2\underline{I}]^{-1} [\underline{y}-\underline{X}_j\hat{\underline{a}}_j] \\
&= \frac{|\underline{X}_j\underline{C}_j\underline{X}_j^t + \sigma_j^2\underline{I} + [\underline{y}-\underline{X}_j\hat{\underline{a}}_j][\underline{y}-\underline{X}_j\hat{\underline{a}}_j]^t|}{|\underline{X}_j\underline{C}_j\underline{X}_j^t + \sigma_j^2\underline{I}|} - 1 \quad (1.12)
\end{aligned}$$

The model posterior probability can then be computed using Bayes theorem

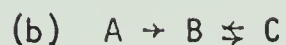
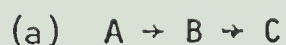
$$p_j'' = \frac{p_j' \cdot L(\underline{y}|\phi_j)}{\sum_{k=1}^m p_k' L(\underline{y}|\phi_k)} \quad (1.13)$$

The expression on the right hand side of (1.12) involves finding the determinant of a matrix of the order equal to the number of data points. When the number of data points is greater than about 20, serious computation errors arise in obtaining the determinants for (1.12). As is often the case in testing equations of state in thermodynamics or certain kinetic models, the number of data points is usually more than 50 and not infrequently as large as 200 or more. To avoid these computation errors and large computing times in obtaining the above determinants, an expression for the expected likelihood, involving only the sufficient statistics and not the entire data would be, indeed, very useful. Such an expression is developed in Chapter 2.

1.3.3 Sequential Design of Experiments

Sometimes, during the study of a physical phenomenon, the data are not sufficient to favour any particular model. In such cases, it becomes necessary to design more experiments which should be carefully planned to discriminate among the rival models. The design of experiments involves finding of a region where some particular model will fit the data very well whereas the other will show a poor fit.

A study frequently reveals the existence of such problems in which data in one region has negligible information while in some other region, it might be substantially more useful in discriminating among the rival models. Box and Hill (1967) have given the example of the following two simple rival mechanisms.



It is required to ascertain one of the above two mechanisms. Figure 1.2 shows the concentration of B against time x_1 . Clearly, the concentrations of B for large times are required to discriminate between the two model forms. Another practical problem in literature is that of the Water Gas Shift Reaction where a large number of mechanisms have been considered [Jenkins, Kisiel, Price and Rippin (1963)].

In discriminating among mechanistic models, divergence in the responses of the various models alone, is not a sufficient criterion for achieving discrimination. It is the divergence in the various responses relative to the limits of errors (say 95% confidence region), that is important. The point of maximum divergence is not necessarily the best discriminating point. As an example, the light lines in Figure 1.3 represent the limits of the errors and the heavy lines indicate the model responses. Obviously, any point in the unshaded region is better for indicating the difference in the models than any other point outside this region, even though greater difference in the responses of the two models exist towards the right of the unshaded region.

The earliest work in model discrimination was that of Williams (1959), Cox (1961, 1962) and, Hunter and Reiner (1965). Williams employed regression analysis, Cox used a hypothesis testing approach and Hunter and Reiner took the designs that maximize the

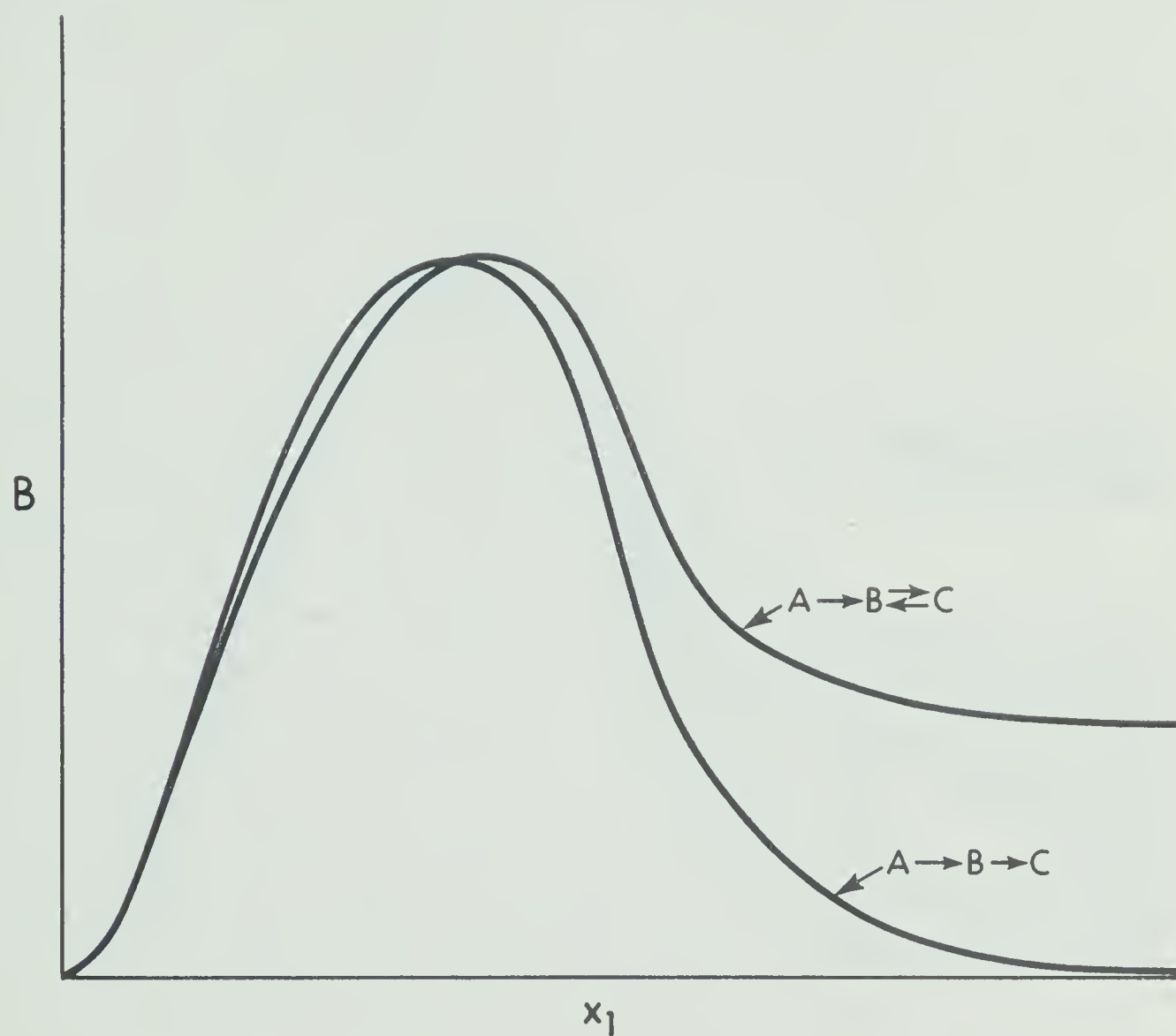


FIGURE 1.2 TWO RIVAL MECHANISMS FOR A CHEMICAL REACTION UNDER INVESTIGATION

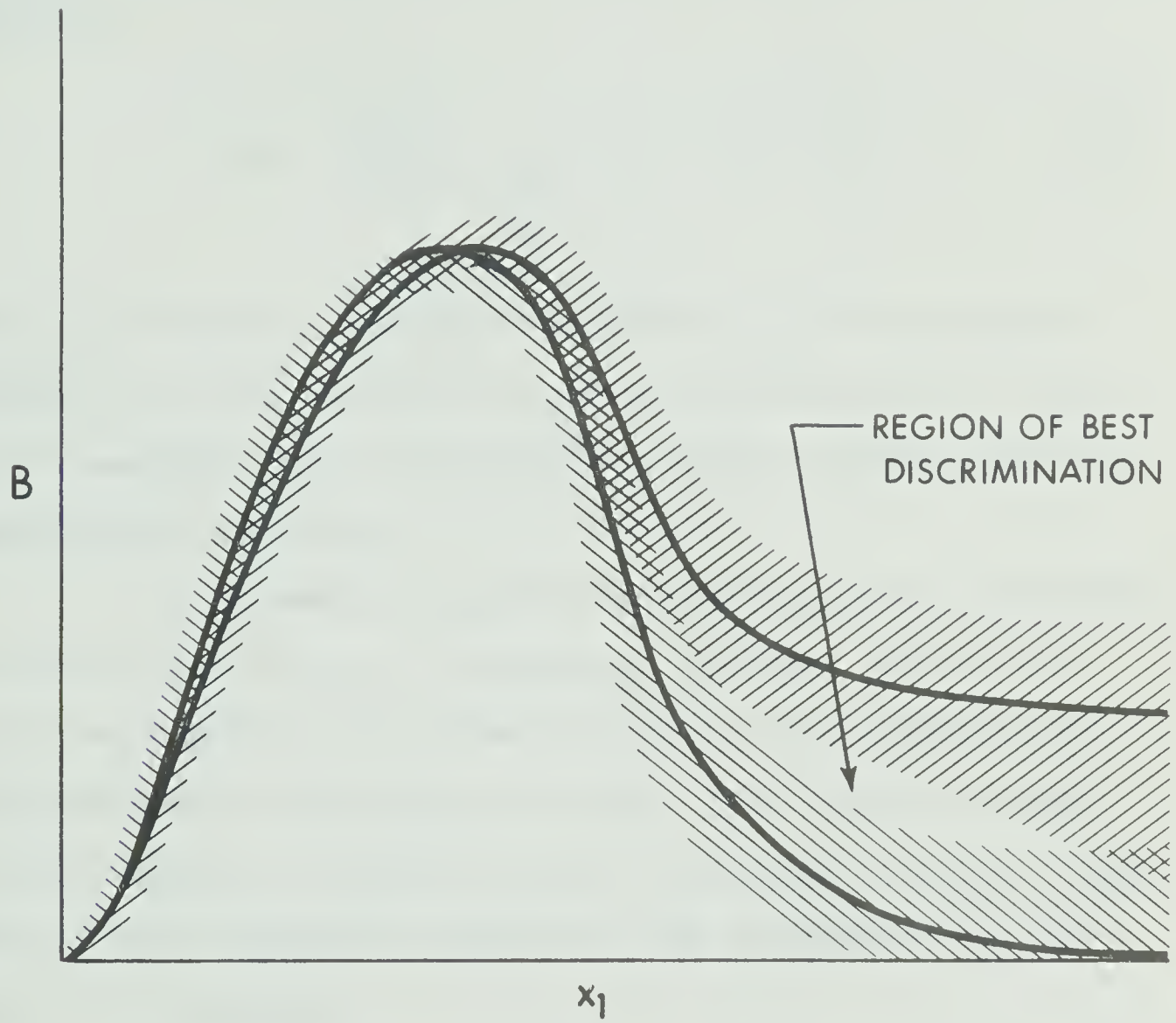


FIGURE 1.3 EFFECT OF ERROR VARIANCE IN MODEL DISCRIMINATION

deviation between the predicted values from the various models. Subsequently, Roth (1965) developed a criterion that calculates the weighted average of the separation from each given model. The weights are the posterior probabilities of the models from all the prior data points. The criterion to choose the $(n+1)$ st experimental point is to maximize

$$Z(x) = \sum_{i=1}^m \left[p_{i,n} \prod_{\substack{i=1 \\ i \neq j}}^m \left| \hat{y}_j(\underline{x}) - \hat{y}_i(\underline{x}) \right| \right] \quad (1.14)$$

where $\underline{x}$ is the set of independent variables, $p_{i,n}$ is the posterior probability of the i th model from n previous observations and $\hat{y}_j(\underline{x})$ is the expected value of the prediction from the j th model under the experimental conditions $\underline{x}$.

All the above criteria fail to take into account uncertainty in the predicted values of the dependent variables. Box and Hill (1967) developed a criterion that accounts for this uncertainty in the prediction of the response of a model. The criterion is based on the concepts of information theory, as developed by Kulback (1959), and the idea of entropy as a measure of the randomness or the uncertainty of a situation.

For a system of m models, the entropy at the end of n observations may be defined as

$$S = - \sum_{i=1}^m p_{i,n} \ln p_{i,n} \quad (1.15)$$

The entropy of a system is at maximum if all the models are equally likely and is zero if one of the models is certainly true.

Box and Hill showed that the negative upper bound of the expected entropy change resulting from the $(n+1)$ st experiment performed at $\underline{x}$ is given by

$$D(\underline{x}) = \frac{1}{2} \sum_{i=1}^m \sum_{j=i+1}^m p_{i,n} p_{j,n} \left\{ \frac{\sigma_i^2 - \sigma_j^2}{(\sigma^2 + \sigma_i^2)(\sigma^2 + \sigma_j^2)} + [\hat{y}_i(\underline{x}) - \hat{y}_j(\underline{x})]^2 \cdot \left[\frac{1}{\sigma^2 + \sigma_i^2} + \frac{1}{\sigma^2 + \sigma_j^2} \right] \right\} \quad (1.16)$$

where σ^2 is the known error variance for a single measurement and $\sigma_i^2 = \underline{x}_D (\underline{X}^T \underline{X})^{-1} \underline{x}_D^T \sigma^2$. $\underline{X}$ is defined in section 1.31 and $\underline{x}_D$ is the row vector that would be the $(n+1)$ st row of $\underline{X}$ for an observation made at $\underline{x}$.

The design of experiments for model discrimination, would now be to choose a point that maximizes D . It will be however, more logical to have a criterion that gives the expected change in entropy rather than an approximation of it in terms of the upper limit as in the case of Box and Hill criterion. Reilly (1969) obtained an expression for the expected change in entropy involving an infinite integral that has to be solved numerically. The expected change in entropy was given to be

$$\begin{aligned}
 R(\underline{x}) = & -\frac{1}{2} (1 + \ln 2\pi) - \sum_{i=1}^m p_{i,n} \left\{ \frac{1}{2} \ln(\sigma^2 + \sigma_i^2) \right. \\
 & \left. + E_{y_{n+1}} [\ln q(\underline{x}, y_{n+1})] \right\}
 \end{aligned} \tag{1.17}$$

where

$$q(\underline{x}, y_{n+1}) = \sum_{i=1}^m p_{i,n} L_i(\underline{x}, y_{n+1})$$

$L_i(\underline{x}, y_{n+1})$ is the expected likelihood of the i th model for the $(n+1)$ st observation and $E_{y_{n+1}} [\ln q(\underline{x}, y_{n+1})]$ is the expected value of $\ln q(\underline{x}, y_{n+1})$ over y_{n+1} where y_{n+1} is normally distributed with mean $\hat{y}_{n+1}(\underline{x})$ and variance $\sigma^2 + \sigma_i^2$. Reilly suggests evaluating the last term numerically using Gauss-Hermite quadrature.

However, it would be convenient to have a simple analytical expression for the expected change in entropy criterion. Such an expression is developed in Chapter 3.

2. EXPECTED LIKELIHOODS OF DATA

2.1 Development of the Expected Likelihood of the Data from the Sufficient Statistics

The inadequacy of the existing expressions to compute the expected likelihood of the given data when a large number of observations is involved was seen in section 1.3.2. Serious computation errors arise when the number of data points is in excess of about 20. In addition, the computing time required to evaluate the determinants of matrices of the order of 100 or more, can be considerable. The difficulty originates from the fact that the entire data set is required in calculating the likelihoods by existing methods. Need for a procedure that would require only the sufficient statistics of the data is apparent. Such expressions are developed in this chapter. In the treatment that follows, errors are assumed to be normally distributed with mean zero and a known variance. Expressions have been developed using both the parameter prior and posterior distributions.

2.1.1 Expected Likelihoods Using Parameter Prior Distribution

When the errors are assumed to be normally distributed with mean zero and variance (known) σ^2 , the joint prior distribution of the parameters may be represented by:

$$f'(\underline{a}) = f_N^{(m)}(\underline{a} | \underline{\mu}_0, \underline{C}_0)$$

$$= (2\pi\sigma^2)^{-m/2} |\underline{C}_0|^{-1/2} \cdot \exp.\left\{-\frac{1}{2} [\underline{a} - \underline{\mu}_0]^t \underline{C}_0^{-1} [\underline{a} - \underline{\mu}_0]\right\} \quad (2.1)$$

where m is the number of parameters, $\underline{C}_0$ and $\underline{\mu}_0$ are respectively the prior variance-covariance matrix and the expected value of $\underline{a}$, $|\underline{C}_0|$ being the determinant of $\underline{C}_0$.

The likelihood of the occurrence of n observations given the j th model is true with the set of parameters $\underline{a}$, is

$$L(\underline{y}|\phi_j, \underline{a}) = (2\pi\sigma^2)^{-n/2} \cdot \exp\left\{-\frac{\hat{S}^2}{2\sigma^2}\right\} \cdot \exp\left\{-\frac{1}{2} [\underline{a}-\hat{\underline{a}}]^t \underline{C}^{-1} [\underline{a}-\hat{\underline{a}}]\right\} \quad (2.2)$$

where $\hat{\underline{a}}$ is the value of the parameter vector $\underline{a}$, that gives the least sum of errors squared, $\hat{S}^2$ and $\underline{C} = \frac{\underline{X}^T \underline{X}}{\sigma^2}$, $\underline{X}$ being the design matrix for the j th model as defined in section 1.3.1.

The expectation of the likelihood of the data over the entire parameter space may now be written as

$$\begin{aligned} L(\underline{y}|\phi_j) &= \int_{-\infty}^{\infty} L(\underline{y}|\phi_j, \underline{a}) \cdot f'(\underline{a}) \cdot d\underline{a} \\ &= \int_{-\infty}^{\infty} (2\pi\sigma^2)^{-n/2} \cdot \exp\left\{-\frac{\hat{S}^2}{2\sigma^2}\right\} \cdot (2\pi)^{-m/2} \cdot |\underline{C}_0|^{-1/2} \\ &\quad \exp\left\{-\frac{1}{2} [\underline{a}-\hat{\underline{a}}]^t \cdot \underline{C}^{-1} [\underline{a}-\hat{\underline{a}}]\right\} \\ &\quad \exp\left\{-\frac{1}{2} [\underline{a}-\underline{\mu}_0]^t \cdot \underline{C}_0^{-1} [\underline{a}-\underline{\mu}_0]\right\} \cdot d\underline{a} \end{aligned} \quad (2.3)$$

Define

$$S = [\underline{a} - \underline{\hat{a}}]^t \underline{C}^{-1} [\underline{a} - \underline{\hat{a}}] + [\underline{a} - \underline{\mu}_O]^t \underline{C}_O^{-1} [\underline{a} - \underline{\mu}_O] \quad (2.4)$$

$$S_1 = [\underline{a} - [\underline{C}_O^{-1} + \underline{C}^{-1}]^{-1} [\underline{C}_O^{-1} \underline{\mu}_O + \underline{C}^{-1} \underline{\hat{a}}]]^t [\underline{C}_O^{-1} + \underline{C}^{-1}] [\underline{a} - [\underline{C}_O^{-1} + \underline{C}^{-1}]^{-1} [\underline{C}_O^{-1} \underline{\mu}_O + \underline{C}^{-1} \underline{\hat{a}}]] \quad (2.5)$$

and

$$S_2 = \underline{\hat{a}}^t \underline{C}^{-1} \underline{\hat{a}} + \underline{\mu}_O^t \underline{C}_O^{-1} \underline{\mu}_O - [\underline{C}_O^{-1} \underline{\mu}_O + \underline{C}^{-1} \underline{\hat{a}}]^t [\underline{C}_O^{-1} + \underline{C}^{-1}]^{-1} [\underline{C}_O^{-1} \underline{\mu}_O + \underline{C}^{-1} \underline{\hat{a}}] \quad (2.6)$$

Now

$$\begin{aligned} S &= [\underline{a} - \underline{\hat{a}}]^t \underline{C}^{-1} [\underline{a} - \underline{\hat{a}}] + [\underline{a} - \underline{\mu}_O]^t \underline{C}_O^{-1} [\underline{a} - \underline{\mu}_O] \\ &= \underline{a}^t \underline{C}^{-1} \underline{a} - \underline{a}^t \underline{C}^{-1} \underline{\hat{a}} - \underline{\hat{a}}^t \underline{C}^{-1} \underline{a} + \underline{\hat{a}}^t \underline{C}^{-1} \underline{\hat{a}} \\ &\quad + \underline{a}^t \underline{C}_O^{-1} \underline{a} - \underline{a}^t \underline{C}_O^{-1} \underline{\mu}_O - \underline{\mu}_O^t \underline{C}_O^{-1} \underline{a} + \underline{\mu}_O^t \underline{C}_O^{-1} \underline{\mu}_O \end{aligned} \quad (2.7)$$

Adding and subtracting $[\underline{C}_O^{-1} \underline{\mu}_O + \underline{C}^{-1} \underline{\hat{a}}]^t [\underline{C}_O^{-1} + \underline{C}^{-1}]^{-1} [\underline{C}_O^{-1} \underline{\mu}_O + \underline{C}^{-1} \underline{\hat{a}}]$

from the right hand side

$$\begin{aligned} S &= S_2 + \{ [\underline{C}_O^{-1} \underline{\mu}_O + \underline{C}^{-1} \underline{\hat{a}}]^t [\underline{C}_O^{-1} + \underline{C}^{-1}]^{-1} [\underline{C}_O^{-1} \underline{\mu}_O + \underline{C}^{-1} \underline{\hat{a}}] \\ &\quad - \underline{a}^t \underline{C}^{-1} \underline{a} - \underline{a}^t \underline{C}^{-1} \underline{\hat{a}} - \underline{\hat{a}}^t \underline{C}^{-1} \underline{a} + \underline{a}^t \underline{C}_O^{-1} \underline{a} - \underline{a}^t \underline{C}_O^{-1} \underline{\mu}_O \\ &\quad - \underline{\mu}_O^t \underline{C}_O^{-1} \underline{a} \} \end{aligned} \quad (2.8)$$

Also

$$\begin{aligned}
S_1 &= [\underline{a}_1 - [\underline{C}_0^{-1} + \underline{C}^{-1}]^{-1}[\underline{C}_0^{-1}\underline{\mu}_0 + \underline{C}^{-1}\hat{\underline{a}}]]^t [\underline{C}_0^{-1} + \underline{C}^{-1}] \\
&\quad [\underline{a} - [\underline{C}_0^{-1} + \underline{C}^{-1}]^{-1}[\underline{C}_0^{-1}\underline{\mu}_0 + \underline{C}^{-1}\hat{\underline{a}}]] \\
&= [\underline{a} - [\underline{C}_0^{-1} + \underline{C}^{-1}]^{-1}[\underline{C}_0^{-1}\underline{\mu}_0 + \underline{C}^{-1}\hat{\underline{a}}]]^t [\underline{C}_0^{-1} + \underline{C}^{-1}] \underline{a} \\
&\quad - [\underline{C}_0^{-1}\underline{\mu}_0 + \underline{C}^{-1}\hat{\underline{a}}]] \\
&= \underline{a}^t [\underline{C}_0^{-1} + \underline{C}^{-1}] \underline{a} - \underline{C}_0^{-1}\underline{\mu}_0 + \underline{C}^{-1}\hat{\underline{a}}]^t \hat{\underline{a}} \\
&\quad + [\underline{C}_0^{-1}\underline{\mu}_0 + \underline{C}^{-1}\hat{\underline{a}}]^t [\underline{C}_0^{-1} + \underline{C}^{-1}]^{-1} [\underline{C}_0^{-1}\underline{\mu}_0 + \underline{C}^{-1}\hat{\underline{a}}] \\
&\quad - \underline{a}^t [\underline{C}_0^{-1}\underline{\mu}_0 + \underline{C}^{-1}\hat{\underline{a}}] \tag{2.9}
\end{aligned}$$

This is the same expression as in the parenthesis of (2.8)

so that now

$$S = S_1 + S_2 \tag{2.10}$$

It may be noted that only S_1 contains the parameter vector as random variables; S_2 is a constant in this respect and may be taken out of the kernel in the expressions of (2.3). Substituting the above results in (2.3) and taking the exponential involving S_2 out of the integral, the expected value of the likelihood now becomes

$$\begin{aligned}
&(2\pi\sigma^2)^{-n/2} \cdot \exp.\left\{-\frac{\hat{S}^2}{2\sigma^2}\right\} \cdot (2\pi)^{-m/2} \cdot |\underline{C}_0|^{-1/2} \cdot \exp.\left\{-\frac{S_2}{2}\right\} \cdot \\
&\int_{-\infty}^{\infty} \exp.\left\{-\frac{1}{2} [\underline{a} - [\underline{C}_0^{-1} + \underline{C}^{-1}]^{-1}[\underline{C}_0^{-1}\underline{\mu}_0 + \underline{C}^{-1}\hat{\underline{a}}]]^t [\underline{C}_0^{-1} + \underline{C}^{-1}] \right. \\
&\quad \left. [\underline{a} - [\underline{C}_0^{-1} + \underline{C}^{-1}]^{-1}[\underline{C}_0^{-1}\underline{\mu}_0 + \underline{C}^{-1}\hat{\underline{a}}]]\right\} \cdot d\underline{a} \tag{2.11}
\end{aligned}$$

But

$$\begin{aligned}
 & (2\pi)^{-m/2} \cdot |\underline{C}_0^{-1} + \underline{C}^{-1}|^{1/2} \cdot \int_{-\infty}^{\infty} \exp. \left\{ -\frac{1}{2} [\underline{a} - [\underline{C}_0^{-1} + \underline{C}^{-1}]^{-1} [\underline{C}_0^{-1} \underline{\mu}_0 + \underline{C}^{-1} \underline{\hat{a}}]]^t \right. \\
 & \quad \left. [\underline{C}_0^{-1} + \underline{C}^{-1}] [\underline{a} - [\underline{C}_0^{-1} + \underline{C}^{-1}]^{-1} [\underline{C}_0^{-1} \underline{\mu}_0 + \underline{C}^{-1} \underline{\hat{a}}]] \right\} \\
 & \quad \cdot d\underline{a} \\
 & = 1 \tag{2.12}
 \end{aligned}$$

Therefore the expected value of the likelihood is

$$\begin{aligned}
 & (2\pi\sigma^2)^{-n/2} \cdot \exp. \left\{ -\frac{\hat{S}^2}{2\sigma^2} \right\} \cdot \frac{|\underline{C}_0|^{-1/2}}{|\underline{C}_0^{-1} + \underline{C}^{-1}|^{-1/2}} \exp. \left\{ -\frac{1}{2} [\underline{\hat{a}}^t \underline{C}^{-1} \underline{\hat{a}} + \underline{\mu}_0^t \underline{C}_0^{-1} \underline{\mu}_0] \right\} \\
 & \quad \exp. \left\{ +\frac{1}{2} [\underline{C}_0^{-1} \underline{\mu}_0 + \underline{C}^{-1} \underline{\hat{a}}]^t [\underline{C}_0^{-1} + \underline{C}^{-1}]^{-1} [\underline{C}_0^{-1} \underline{\mu}_0 + \underline{C}^{-1} \underline{\hat{a}}] \right\} \\
 & \tag{2.13}
 \end{aligned}$$

The above expression reduces to zero for a uniform improper prior.

2.1.2 Expected Likelihoods Using the Parameter Posterior Distribution

The posterior distribution of parameters is given by

$$\begin{aligned}
 f''(\underline{a}) &= f_N^{(m)}(\underline{a} | [\underline{C}_O^{-1} + \underline{C}^{-1}]^{-1} [\underline{C}_O^{-1} \underline{\mu}_O + \underline{C}^{-1} \underline{\hat{a}}], [\underline{C}_O^{-1} + \underline{C}^{-1}]^{-1}) \\
 &= (2\pi\sigma^2)^{-m/2} \cdot |\underline{C}_O^{-1} + \underline{C}^{-1}|^{1/2} \cdot \\
 &\quad \exp.\{-\frac{1}{2} [\underline{a} - [\underline{C}_O^{-1} + \underline{C}^{-1}]^{-1} [\underline{C}_O^{-1} \underline{\mu}_O + \underline{C}^{-1} \underline{\hat{a}}]]^t [\underline{C}_O^{-1} + \underline{C}^{-1}] \\
 &\quad [\underline{a} - [\underline{C}_O^{-1} + \underline{C}^{-1}]^{-1} [\underline{C}_O^{-1} \underline{\mu}_O + \underline{C}^{-1} \underline{\hat{a}}]]\} \quad (2.14)
 \end{aligned}$$

The expected likelihood of the data for the given model using the posterior joint distribution of parameter is

$$\begin{aligned}
 &\int_{-\infty}^{\infty} (2\pi\sigma^2)^{-n/2} \cdot \exp.\{-\frac{\hat{S}^2}{2\sigma^2}\} \cdot (2\pi)^{-m/2} \cdot |\underline{C}_O^{-1} + \underline{C}^{-1}|^{1/2} \cdot \\
 &\quad \exp.\{-\frac{1}{2} [\underline{a} - \underline{\hat{a}}]^t \underline{C}^{-1} [\underline{a} - \underline{\hat{a}}]\} \\
 &\quad \exp.\{-\frac{1}{2} [\underline{a} - [\underline{C}_O^{-1} + \underline{C}^{-1}]^{-1} [\underline{C}_O^{-1} \underline{\mu}_O + \underline{C}^{-1} \underline{\hat{a}}]]^t [\underline{C}_O^{-1} + \underline{C}^{-1}] \\
 &\quad [\underline{a} - [\underline{C}_O^{-1} + \underline{C}^{-1}]^{-1} [\underline{C}_O^{-1} \underline{\mu}_O + \underline{C}^{-1} \underline{\hat{a}}]] \cdot d\underline{a} \quad (2.15)
 \end{aligned}$$

Let

$$\begin{aligned}
 s &= [\underline{a} - [\underline{C}_O^{-1} + \underline{C}^{-1}]^{-1} [\underline{C}_O^{-1} \underline{\mu}_O + \underline{C}^{-1} \underline{\hat{a}}]]^t [\underline{C}_O^{-1} + \underline{C}^{-1}] \\
 &\quad [\underline{a} - [\underline{C}_O^{-1} + \underline{C}^{-1}]^{-1} [\underline{C}_O^{-1} \underline{\mu}_O + \underline{C}^{-1} \underline{\hat{a}}]] \quad (2.16)
 \end{aligned}$$

$$\begin{aligned}
s_1 &= [\underline{a} - [\underline{C}_O^{-1} + 2\underline{C}^{-1}]^{-1} [\underline{C}_O^{-1} \underline{\mu}_O + 2\underline{C}^{-1} \underline{\hat{a}}]]^t [\underline{C}_O^{-1} + 2\underline{C}^{-1}] \\
&\quad [\underline{a} - [\underline{C}_O^{-1} + 2\underline{C}^{-1}]^{-1} [\underline{C}_O^{-1} \underline{\mu}_O + 2\underline{C}^{-1} \underline{\hat{a}}]] \quad (2.17)
\end{aligned}$$

Expanding the right hand side of (2.17) and regrouping

$$\begin{aligned}
s_1 &= s + \underline{a}^t \underline{C}^{-1} \underline{a} - \underline{\hat{a}}^t \underline{C}^{-1} \underline{a} - \underline{a}^t \underline{C}^{-1} \underline{\hat{a}} \\
&\quad + [\underline{C}_O^{-1} \underline{\mu}_O + 2\underline{C}^{-1} \underline{\hat{a}}]^t [\underline{C}_O^{-1} + 2\underline{C}^{-1}]^{-1} [\underline{C}_O^{-1} \underline{\mu}_O + 2\underline{C}^{-1} \underline{\hat{a}}] \\
&\quad - [\underline{C}_O^{-1} \underline{\mu}_O + \underline{C}^{-1} \underline{\hat{a}}]^t [\underline{C}_O^{-1} + \underline{C}^{-1}]^{-1} [\underline{C}_O^{-1} \underline{\mu}_O + \underline{C}^{-1} \underline{\hat{a}}]
\end{aligned}$$

or

$$\begin{aligned}
&s + [\underline{a} - \underline{\hat{a}}]^t [\underline{C}^{-1}] [\underline{a} - \underline{\hat{a}}] \\
&= s_1 - [\underline{C}_O^{-1} \underline{\mu}_O + 2\underline{C}^{-1} \underline{\hat{a}}] [\underline{C}_O^{-1} + 2\underline{C}^{-1}]^{-1} [\underline{C}_O^{-1} \underline{\mu}_O + 2\underline{C}^{-1} \underline{\hat{a}}] \\
&\quad + [\underline{C}_O^{-1} \underline{\mu}_O + \underline{C}^{-1} \underline{\hat{a}}]^t [\underline{C}_O^{-1} + \underline{C}^{-1}]^{-1} [\underline{C}_O^{-1} \underline{\mu}_O + \underline{C}^{-1} \underline{\hat{a}}] \\
&\quad + \underline{\hat{a}}^t \underline{C}^{-1} \underline{\hat{a}} \quad (2.18)
\end{aligned}$$

The expected likelihood may now be written as

$$\begin{aligned}
 & (2\pi\sigma^2)^{-n/2} \cdot \exp.\left\{-\frac{\hat{S}^2}{2\sigma^2}\right\} \cdot (2\pi)^{-m/2} \cdot |\underline{C}_O^{-1} + \underline{C}^{-1}|^{1/2} \cdot \exp.\left\{-\frac{1}{2} \underline{\hat{a}}^t \underline{C}^{-1} \underline{\hat{a}}\right\} \cdot \\
 & \exp.\left\{-\frac{1}{2} [(\underline{C}_O^{-1} \underline{\mu}_O + \underline{C}^{-1} \underline{\hat{a}})^t (\underline{C}_O^{-1} + \underline{C}^{-1})^{-1} (\underline{C}_O^{-1} \underline{\mu}_O + \underline{C}^{-1} \underline{\hat{a}}) \right. \\
 & \quad \left. - (\underline{C}_O^{-1} \underline{\mu}_O + 2\underline{C}^{-1} \underline{\hat{a}})^t (\underline{C}_O^{-1} + 2\underline{C}^{-1})^{-1} (\underline{C}_O^{-1} \underline{\mu}_O + 2\underline{C}^{-1} \underline{\hat{a}})]\right\} \cdot \\
 & \int_{-\infty}^{\infty} \exp.\left\{-\frac{1}{2} [\underline{a} - (\underline{C}_O^{-1} + 2\underline{C}^{-1})^{-1} (\underline{C}_O^{-1} \underline{\mu}_O + 2\underline{C}^{-1} \underline{\hat{a}})]^t (\underline{C}_O^{-1} + 2\underline{C}^{-1}) \right. \\
 & \quad \left. [\underline{a} - (\underline{C}_O^{-1} + 2\underline{C}^{-1})^{-1} (\underline{C}_O^{-1} \underline{\mu}_O + 2\underline{C}^{-1} \underline{\hat{a}})]\right\} \cdot d\underline{a}
 \end{aligned}$$

But

$$\begin{aligned}
 & (2\pi)^{-m/2} \cdot |\underline{C}_O^{-1} + 2\underline{C}^{-1}|^{1/2} \cdot \int_{-\infty}^{\infty} \exp.\left\{-\frac{1}{2} [\underline{a} - (\underline{C}_O^{-1} + 2\underline{C}^{-1})^{-1} (\underline{C}_O^{-1} \underline{\mu}_O + 2\underline{C}^{-1} \underline{\hat{a}})]^t \right. \\
 & \quad \left. (\underline{C}_O^{-1} + 2\underline{C}^{-1}) [\underline{a} - (\underline{C}_O^{-1} + 2\underline{C}^{-1})^{-1} (\underline{C}_O^{-1} \underline{\mu}_O + 2\underline{C}^{-1} \underline{\hat{a}})]\right\} \cdot d\underline{a} \\
 & = 1
 \end{aligned} \tag{2.19}$$

Therefore the expected likelihood for the given model is

$$\begin{aligned}
 & (2\pi\sigma^2)^{-n/2} \cdot \exp.\left\{-\frac{\hat{S}^2}{2\sigma^2}\right\} \cdot \frac{|\underline{C}_O^{-1} + \underline{C}^{-1}|^{1/2}}{|\underline{C}_O^{-1} + 2\underline{C}^{-1}|^{1/2}} \cdot \exp.\left\{-\frac{1}{2} \underline{\hat{a}}^t \underline{C}^{-1} \underline{\hat{a}}\right\} \cdot \\
 & \exp.\left\{-\frac{1}{2} [(\underline{C}_O^{-1} \underline{\mu}_O + \underline{C}^{-1} \underline{\hat{a}})^t (\underline{C}_O^{-1} + \underline{C}^{-1})^{-1} (\underline{C}_O^{-1} \underline{\mu}_O + \underline{C}^{-1} \underline{\hat{a}}) \right. \\
 & \quad \left. - (\underline{C}_O^{-1} \underline{\mu}_O + 2\underline{C}^{-1} \underline{\hat{a}})^t (\underline{C}_O^{-1} + 2\underline{C}^{-1})^{-1} (\underline{C}_O^{-1} \underline{\mu}_O + 2\underline{C}^{-1} \underline{\hat{a}})]\right\} \tag{2.20}
 \end{aligned}$$

For uniform improper prior $\underline{C}_0^{-1} = \underline{0}$ and the above expression reduces to

$$(2\pi\sigma^2)^{-n/2} \cdot \exp. \left\{ -\frac{\hat{S}^2}{2\sigma^2} \right\} \cdot \left(\frac{1}{2}\right)^{m/2} \quad (2.21)$$

This simple expression gives a theoretical justification for the conventional least squares criterion. When the number of parameters m , is the same for all the competing models, $\hat{S}^2$ is a quantity that in itself uniquely defines the model performance. The likelihood is only a function of the $\hat{S}^2$ of the model. In general, this quantity should be corrected by an additional factor of $(\frac{1}{2})^{m/2}$, so as to favour the model with fewer parameters. This is evident from (2.21)

3. SEQUENTIAL DESIGN OF EXPERIMENTS FOR MODEL DISCRIMINATION

3.1 Development of Expected Entropy Change Criterion for Model Discrimination

The objective here is to obtain a criterion that will choose the experimental condition $\underline{x}$, such that the response of the system will be expected to result in maximum discrimination between the competing models. We start with the concept of entropy as the measure of uncertainty associated with the knowledge of a system. Shannon (1948) introduced this concept of entropy in communication problems. For a set of m competing models, the negative of expected entropy change (or the expected rate of transmission of information) for the $(n+1)$ st observation is

$-R$ = entropy at the input - expected entropy at the output

$$= - \sum_{i=1}^m p_{n,i} \ln p_{n,i} - (-1) \int_{y_{n+1}} \left[\sum_{i=1}^m p_{n+1,i} \ln p_{n+1,i} \right] f(y_{n+1}) dy_{n+1} \quad (3.1)$$

where $f(y_{n+1})$ is the unconditional distribution of the prediction y_{n+1} , and $p_{n+1,i}$ is the resultant probability of the i th model after the $(n+1)$ st observation has been utilized.

After observing y_{n+1} , the posterior model probabilities $p_{n+1,i}$ can be computed from Bayes' theorem

$$p_{n+1,i} = \frac{p_{n,i} L(y_{n+1} | \phi_i)}{\sum_{j=1}^m p_{n,j} L(y_{n+1} | \phi_j)} \quad (3.2)$$

where $L(y_{n+1} | \phi_j)$, the expected likelihood of observing y_{n+1} given the model form ϕ_j is true, is obtained by integrating the likelihood $L(y_{n+1} | \phi_j, \underline{a}_j)$ of observing y_{n+1} if ϕ_j is true with parameters $\underline{a}_j$ over the entire parameter space $\underline{a}_j$, that is

$$L(y_{n+1} | \phi_j) = \int_{-\infty}^{\infty} L(y_{n+1} | \phi_j, \underline{a}_j) f(\underline{a}_j) d\underline{a}_j \quad (3.3)$$

and $f(\underline{a}_j)$ being the prior parameter distribution which is multivariate normal and incorporates the n prior data points. From the assumption that the errors are normally distributed with known error variance σ^2

$$L(y_{n+1} | \phi_j, \underline{a}_j) = f_N(y_{n+1} | \hat{y}_{n+1,j}, \sigma^2) \quad (3.4)$$

where $\hat{y}_{n+1,j}$ is the expected value of the prediction from the j th model at the value of $\underline{x}_{n+1,j}$ corresponding to the experimental condition where y_{n+1} has been observed. Substituting (3.4) into (3.3) leads to

$$L(y_{n+1} | \phi_j) = f_N(y_{n+1} | \hat{y}_{n+1,j}, \sigma^2 + \sigma_{n+1,j}^2) \quad (3.5)$$

where

$$\sigma_{n+1,j}^2 = \underline{x}_{n+1,j} (\underline{X}_{n,j}^t \underline{X}_{n,j})^{-1} \underline{x}_{n+1,j}^t \sigma^2 \quad (3.6)$$

as defined in section 1.3.3 the term σ^2 accounts for the error due to a single measurement and $\sigma_{n+1,j}^2$ describes the error due to the uncertainty in the parameters given that the model form ϕ_j is correct; $\underline{X}_{n,j}$ is the design matrix of the n prior observations and $\underline{x}_{n+1,j}$ is the design vector for the j th model under the experimental conditions for observing y_{n+1} , as defined in sections 1.3.1 and 1.3.3.

After y_{n+1} has been observed, all the quantities $L(y_{n+1}|\phi_i)$ and $p_{n+1,i}$ are scalars. However, before the $(n+1)$ st observation has actually been made, we treat y_{n+1} as a random variable and $L(y_{n+1}|\phi_i)$ and $p_{n+1,i}$ in (3.2) are now functions of a random variable and are themselves random variables. Furthermore, it is assumed that the set of models under consideration, which we call a metamodel, is mutually exhaustive, i.e. the data is generated by a model within the metamodel. Of course, in many situations, it is always possible that some other model may describe the data better than all the models in metamodel. However, we cannot take into account that which we do not perceive. Given this, it follows that

$$f(y_{n+1}) = \sum_{j=1}^m p_{n,j} L(y_{n+1}|\phi_j) \quad (3.7)$$

that is, the unconditional density of y_{n+1} is a weighted average of the likelihoods of generating the data from the various models under consideration. This assumption enables the exact solution of the integral in (3.1)

Substituting (3.7) and (3.2) in (3.1)

$$-R = - \sum_{i=1}^m p_{n,i} \ln p_{n,i} + \int_{y_{n+1}} \left[\sum_{i=1}^m p_{n,i} L(y_{n+1}|\phi_i) \cdot \ln \frac{p_{n,i} L(y_{n+1}|\phi_i)}{f(y_{n+1})} \right] dy_{n+1} \quad (3.8)$$

The integral in (3.8) may be expanded to give

$$\int_{y_{n+1}} \left[\sum_{i=1}^m p_{n,i} L(y_{n+1} | \phi_i) \ln p_{n,i} + \sum_{i=1}^m p_{n,i} L(y_{n+1} | \phi_i) \ln \frac{L(y_{n+1} | \phi_i)}{f(y_{n+1})} \right] dy_{n+1} \quad (3.9)$$

Since $\int_{y_{n+1}} L(y_{n+1} | \phi_i) dy_{n+1} = 1$, (3.9) becomes

$$\sum_{i=1}^m p_{n,i} \ln p_{n,i} + \int_{y_{n+1}} \left[\sum_{i=1}^m p_{n,i} L(y_{n+1} | \phi_i) \ln \frac{L(y_{n+1} | \phi_i)}{f(y_{n+1})} \right] dy_{n+1} \quad (3.10)$$

Substituting in (3.8)

$$-R = \sum_{i=1}^m p_{n,i} \left[\int_{y_{n+1}} L(y_{n+1} | \phi_i) \ln \frac{L(y_{n+1} | \phi_i)}{f(y_{n+1})} dy_{n+1} \right] \quad (3.11)$$

From (3.6) and (3.7), the unconditional distribution of the dependent variable y_{n+1} may be written in terms of $\hat{y}_{n+1}$ and σ_{n+1}^2 where

$$\hat{y}_{n+1} = \sum_{i=1}^m p_{n,i} \hat{y}_{n+1,i} \quad (3.12)$$

and

$$\sigma_{n+1}^2 = \sum_{i=1}^m p_{n,i} \sigma_{n+1,i}^2 \quad (3.13)$$

So that

$$f(y_{n+1}) = f_N(y_{n+1} | \hat{y}_{n+1}, \sigma^2 + \sigma_{n+1}^2) \quad (3.14)$$

Utilizing the above relations, the integral in (3.11) reduces to

$$\begin{aligned}
 & \int_{y_{n+1}} L(y_{n+1} | \phi_i) \ln \frac{L(y_{n+1} | \phi_i)}{f(y_{n+1})} dy_{n+1} \\
 &= \int_{y_{n+1}} L(y_{n+1} | \phi_i) \ln L(y_{n+1} | \phi_i) dy_{n+1} \\
 &- \int_{y_{n+1}} L(y_{n+1} | \phi_i) \ln f(y_{n+1} | \phi_o) dy_{n+1} \quad (3.15)
 \end{aligned}$$

The first integral on the right hand side of (3.15) can be evaluated as follows

$$\int_{y_{n+1}} L(y_{n+1} | \phi_i) \left\{ -\frac{1}{2} \ln 2\pi(\sigma^2 + \sigma_{n+1,i}^2) - \frac{1}{2} \frac{(y_{n+1} - \hat{y}_{n+1,i})^2}{(\sigma^2 + \sigma_{n+1,i}^2)} \right\} dy_{n+1}$$

The expected value of $(y_{n+1} - \hat{y}_{n+1,i})^2$ with respect to y_{n+1} , assuming model i to be correct, is $(\sigma^2 + \sigma_{n+1,i}^2)$. Consequently, the above integral becomes

$$-\frac{1}{2} [1 + \ln 2\pi(\sigma^2 + \sigma_{n+1,i}^2)] \quad (3.16)$$

The second integral on the right of (3.15) can be written as

$$\int_{y_{n+1}} L(y_{n+1} | \phi_i) \left\{ -\frac{1}{2} \ln 2\pi(\sigma^2 + \sigma_{n+1}^2) - \frac{1}{2} \frac{(y_{n+1} - \hat{y}_{n+1})^2}{\sigma^2 + \sigma_{n+1}^2} \right\} dy_{n+1} \quad (3.17)$$

$$\begin{aligned} \text{Also, } (y_{n+1} - \hat{y}_{n+1})^2 &= (y_{n+1} - \hat{y}_{n+1,i})^2 + (\hat{y}_{n+1,i} - \hat{y}_{n+1})^2 \\ &+ 2(y_{n+1} - \hat{y}_{n+1,i})(y_{n+1} - \hat{y}_{n+1}) \end{aligned} \quad (3.18)$$

The expected value of the last term in (3.16) is zero and the second term is just a constant. (3.17) now simplifies to

$$- \frac{1}{2} \left\{ \ln 2\pi(\sigma^2 + \sigma_{n+1}^2) + \frac{\sigma^2 + \sigma_{n+1,i}^2}{\sigma^2 + \sigma_{n+1}^2} + \frac{(\hat{y}_{n+1,i} - \hat{y}_{n+1})^2}{\sigma^2 + \sigma_{n+1}^2} \right\} \quad (3.19)$$

Substituting the results of (3.16) and (3.19) in (3.8)

$$- R = 0.5 \sum_{i=1}^m p_{n,i} \left[- \ln \frac{\sigma^2 + \sigma_{n+1,i}^2}{\sigma^2 + \sigma_{n+1}^2} + \frac{\sigma_{n+1,i}^2 - \sigma_{n+1}^2}{\sigma^2 + \sigma_{n+1}^2} + \frac{(\hat{y}_{n+1,i} - \hat{y}_{n+1})^2}{\sigma^2 + \sigma_{n+1}^2} \right] \quad (3.20)$$

Now, the second term of (3.20) disappears as can be seen from the relation (3.13). Thus

$$- R = 0.5 \sum_{i=1}^m p_{n,i} \left[- \ln \frac{\sigma^2 + \sigma_{n+1,i}^2}{\sigma^2 + \sigma_{n+1}^2} + \frac{(\hat{y}_{n+1,i} - \hat{y}_{n+1})^2}{\sigma^2 + \sigma_{n+1}^2} \right] \quad (3.21)$$

The expression appeals in its simplicity. It also supports the attempt of earlier workers to maximize the difference in the response of various models by the inclusion of $(\hat{y}_{n+1,i} - \hat{y}_{n+1})^2$ in (3.21) but restrains it with the uncertainty term $(\sigma^2 + \sigma_{n+1}^2)$. It is the ratio of these two terms that is important. In most cases studied, the first term of (3.21) was significantly less important.

When one model is known for certain, $\sigma_{n+1}^2 = \sigma_{n+1,i}^2$ and $\hat{y}_{n+1} = \hat{y}_{n+1,i}$, where the i th model is the correct one. Also $p_{n,j} = 0$ $j \neq i$, this gives

$$R = 0$$

which is expected.

For comparison, a retrospective view of the two most important earlier design criteria due to Roth and Box and Hill may be provided here. Consistent to the notation used here, Roth's criterion is to choose the $(n+1)$ st experimental setting that maximizes

$$Z(\underline{x}) = \sum_{i=1}^m \left[p_{n,i} \sum_{\substack{j=1 \\ j \neq i}}^m \pi \left| \hat{y}_{n+1,j}(\underline{x}) - y_{n+1,i}(\underline{x}) \right| \right]$$

Box and Hill required maximizing the negative upper bound of the expected entropy change

$$D(\underline{x}) = \frac{1}{2} \sum_{i=1}^m \sum_{j=i+1}^m p_{n,i} p_{n,j} \left\{ \frac{\sigma_{n+1,i}^2 - \sigma_{n+1,j}^2}{(\sigma^2 + \sigma_{n+1,i}^2)(\sigma^2 + \sigma_{n+1,j}^2)} + \left[y_{n+1,i}(\underline{x}) - y_{n+1,j}(\underline{x}) \right] \left[\frac{1}{\sigma^2 + \sigma_{n+1,i}^2} + \frac{1}{\sigma^2 + \sigma_{n+1,j}^2} \right] \right\}$$

It may be noted that Roth's criterion failed to take into account the uncertainty in the response and was based on more or less an empirical interpretation of model discrimination. Box and Hill's criterion was developed from the concept of expected entropy change but is not exact. An upper bound of the theoretically

desirable 'expected entropy change' is used as an index for the design of experiments. The proposed criterion in (1.21) is both simple and exact without involving any further restrictions.

3.2 Examples

The examples chosen to illustrate the use of expected change in entropy criterion, have all appeared in literature [Hill (1966)]. The first two examples involve linear models and the last one consists of discriminating between non-linear kinetic models.

Example 3.1

Consider discriminating between the following polynomial models.

$$\text{Model 1.} \quad y^{(1)} = a_1^{(1)} x$$

$$\text{Model 2.} \quad y^{(2)} = a_1^{(2)} + a_2^{(2)} x$$

$$\text{Model 3.} \quad y^{(3)} = a_1^{(3)} + a_2^{(3)} x + a_3^{(3)} x^2$$

$$\text{Model 4.} \quad y^{(4)} = a_1^{(4)} x + a_2^{(4)} x^2$$

The above four models represent respectively, a straight line through the origin, a straight line not necessarily through the origin, a general quadratic and a quadratic forced through the origin. The data was generated from model 3 with $a_1^{(3)} = a_2^{(3)} = a_3^{(3)} = 1$ and error variance of $\sigma^2 = 1$. The independent variable was constrained to $0 \leq x \leq 10$ and integer values. Starting with equal probabilities for all the models, that is $p_{0,1} = p_{0,2} = p_{0,3} = p_{0,4} = .25$ and

using five data points at $x = 0, 1, 2, 3$ and 4 , the model probabilities were $p_{s,1} = 2.5 \times 10^{-5}$, $p_{s,2} = 6.9 \times 10^{-5}$, $p_{s,3} = .445$ and $p_{s,4} = .554$. The experimental space was then scanned to find out the value of the independent variable that gives maximum expected entropy change. The results are given in Table 3.1. Actual change in entropy was calculated to compare with the expected change in entropy. The results given here are typical of various runs. Total actual entropy change was generally found close to the total expected change in entropy for the entire run.

In this particular example, the optimum experimental point was always found to be zero, which is clearly expected from the model forms of the two important competing models, that is model 3 and model 4. The posterior density of $\underline{a}$ was used in computing the model probabilities.

Example 3.2

This example is to illustrate the model discrimination procedure when some more complex models are reducible to the true model form.

Table 3.1
 Results When Discriminating Among
 the Four Polynomial Models
 (Model 3 True)

n	x	$p_{n,1}$	$p_{n,2}$	$p_{n,3}$	$p_{n,4}$	Expected Entropy Change -R	Actual Entropy Change
0		.25	.25	.25	.25		
5		-	.001	.445	.554		
6	0	-	.001	.881	.118	.151	.272
7	0	-	.001	.963	.036	.272	.209
8	0	-	.001	.948	.051	.202	-.048
9	0	-	.001	.943	.056	.037	-.013
10	0	-	.001	.942	.057	.029	-.003
11	0	-	.001	.962	.037	.024	.062
12	0	-	.001	.997	.002	.016	.142

Taking the same set of models as in Example 3.1, it is clear that model 3 is reducible to model 2 by putting $a_3^{(3)} = 0$. Thus if the data were generated by model 2, model 3 will fit the data equally well and it might be thought impossible to discriminate between these two model forms. In fact, the procedure does discriminate between the two models to favour the model with the fewer number of parameters, the reason being that the average variance of the prediction from the model with the least number of parameters is smallest. It is also evident from our new expressions of the expected likelihood in Chapter 2, where a factor of $(\frac{1}{2})^{\frac{m}{2}}$ gives the maximum expected likelihood for the model with least m , when all such models have the same $\hat{S}^2$.

The data, in this example, was generated from model 2 with $a_1^{(2)} = a_2^{(2)} = 1$ and unit error variance. To start with, all models were considered equally likely and first 5 points data gave the posterior model probabilities of $p_{5,1} = .32$, $p_{5,2} = .251$, $p_{5,3} = .187$ and $p_{5,4} = .233$. The results are shown in Table 3.2. There is a marked preference for the simpler model (model 2) after 12 additional data points, although more runs are required for further discrimination.

Table 3.2

Results When Discriminating Among the Four Polynomial Models

(Model 2 True, Model 3 is Reducible to Model 2)

n	x	$P_{n,1}$	$P_{n,2}$	$P_{n,3}$	$P_{n,4}$	Expected Entropy Change	Actual Entropy Change
0		.25	.25	.25	.25	-	
5		.329	.251	.187	.233	-	
6	0	.179	.300	.163	.358	.70	.033
7	0	.304	.377	.109	.209	.13	.033
8	0	.038	.742	.194	.026	.113	.540
9	0	.003	.775	.221	.002	.052	.20
10	0	.002	.759	.237	.002	.005	-.016
11	0	.003	.737	.258	.002	.004	-.029
12	0	-	.733	.267	-	.005	.022
13	4	-	.745	.255	-	.001	.012
14	10	-	.779	.221	-	.001	.039
15	4	-	.778	.222	-	.004	-.001
16	4	-	.750	.249	-	.004	-.032
17	4	-	.805	.195	-	.006	.068

Example 3.3

Suppose an investigator is studying a set of non-linear models, it could be, say, the four different mechanisms of the reaction $A \rightarrow B$, depending upon whether the reaction is first, second, third or fourth order. The corresponding models to describe these mechanisms shall be

$$\text{Model 1.} \quad y^{(1)} = \exp.\{ - a_1^{(1)} x_1 \exp(- a_2^{(1)}/x_2)\}$$

$$\text{Model 2.} \quad y^{(2)} = \{1 + a_1^{(2)} \exp.(- a_2^{(2)}/x_2)\}^{-1}$$

$$\text{Model 3.} \quad y^{(3)} = \{1 + 2a_1^{(3)} \exp.(- a_2^{(3)}/x_2)\}^{-1/2}$$

$$\text{Model 4.} \quad y^{(4)} = \{1 + 3a_1^{(3)} \exp.(- a_2^{(4)}/x_2)\}^{-1/3}$$

where x_1 is the reaction time, x_2 is the temperature, $y^{(i)}$ is the concentration of the reactant and $a_1^{(i)}$ and $a_2^{(i)}$ are the parameters of the models. It is desired to determine the most likely model by the Bayesian model discrimination procedure. The independent variables are constrained by the process to

$$0 \leq x_1 \leq 150 \quad \text{and}$$

$$450 \leq x_2 \leq 600$$

A grid spacing of 25 is suggested for scanning the experimental space. The parameter values are known to be between $300 \leq a_1^{(i)} \leq 400$ and $5000 \leq a_2^{(i)} \leq 6000$ for all $i = 1, 2, 3$ and 4. Variance of the experimental measurements σ^2 is .0025. In this example, two cases were investigated; (a) model 2 is correct, (b) model 3 is correct.

Both times, data was generated with $a_1^{(i)} = 350$ and $a_2^{(i)} = 5500$. All models were taken to be equally likely at the start and a prior parameter variance-covariance matrix was taken to be

$$\begin{bmatrix} 10^4 & 0 \\ 0 & 10^6 \end{bmatrix}$$

A scheme due to Marquardt (1963) was utilized to obtain $\hat{\underline{a}}$, the least square estimators of the parameters.

Table 3.3 and 3.4 show the results of cases (a) and (b) respectively. The discrimination was very rapid and favoured decisively, the right models within 3 observations. In both the cases, the exploration for the first experimental point gives maximum expected entropy change at (100, 600) and thereafter the cornermost point (150, 600) is favoured. To illustrate the extent of discrimination in this example, Table 3.5 gives the expected entropy change for the various grid points for the search of second point in case (b). The results show sharp discrimination at the higher values of x_1 and x_2 and no discrimination at all when $x_1 = 0$. In another example (model 2 true) 40 data points were generated at (150,450) and the final results were found to be $p_{40,1} = .56$, $p_{40,2} = .41$, $p_{40,3} = .03$ and $p_{40,4} = 0$. This illustrates the tremendous difference in the information contained in the experiments that are carried out at (150, 450) and those at (150, 600). The utility of the correct experimental design is obvious in such situations.

By comparing the results obtained by Hill (1966) for this example, it was found that the design point given by the Box and Hill criterion and that given by the expected entropy change criterion proposed here, are not always the same.

Table 3.3

Results When Discriminating Between the
Four Kinetic Models (Model 2 True)

n	x_1	x_2	$p_{n,1}$	$p_{n,2}$	$p_{n,3}$	$p_{n,4}$
0			.25	.25	.25	.25
1	100	600	.001	.588	.390	.021
2	150	600	-	.588	.390	-
3	150	600	-	.999	.001	-
4	150	600	-	1.000	-	-

Table 3.4

Results When Discriminating Between the
Four Kinetic Models (Model 3 True)

n	x_1	x_2	$p_{n,1}$	$p_{n,2}$	$p_{n,3}$	$p_{n,4}$
0			.25	.25	.25	.25
1	100	600	10^{-6}	.044	.449	.508
2	150	600	-	.078	.855	.067
3	150	600	-	.007	.989	.003

Table 3.5
 Search for the Best Experimental Conditions
 While Discriminating Between the Four Kinetic
 Models (Model 3 True) Expected Change in
 Entropy at Various Operable Conditions

x_1	x_2	Expected Change in Entropy
0	450	.0
50	450	.008
150	450	.018
0	500	.0
150	500	.021
0	550	.0
150	550	.207
0	600	.0
50	600	.156
100	600	.350
150	600	.451

4. STANDARDIZATION OF THE BAYESIAN MODEL DISCRIMINATION PROCEDURE

4.1 Examination of the Bayesian Procedure

The Bayesian method of discriminating among the rival models by using the expected likelihood of the data to compute the posterior model probabilities has been recognized as the best available technique for model discrimination. By using the expected likelihood rather than the maximum likelihood as the criterion for discrimination, any asymmetry in the likelihood function is accounted for. Furthermore, unstructured information in the form of prior beliefs about the model form or the parameters can be utilized by assigning a prior distribution to the parameters of each competing model [see Quon (1970)].

However, some areas of the Bayesian procedure are not well defined and are controversial. This chapter is written with a view of rationalizing the Bayesian model discrimination procedure to some extent.

A few points worth analysing are: (a) assumptions associated with the errors and estimates of error variance used in computing the likelihoods, (b) whether prior distribution or the posterior distribution of parameters should be used in computing the expected likelihoods, (c) will a given amount of information containing various sets of data, always lead to the same level of discrimination, regardless of the order or the groupings of data, when used piecemeal to discriminate the models sequentially. These problems are investigated here and a rationalized procedure is recommended in the next section.

4.1.1 Errors and Error Variance

Suppose the model under consideration is

$$y = \phi(\underline{x}, \underline{a}) + \varepsilon$$

where ε represents the error between the actual observed values of the dependent variable and the one predicted by the model. This error term, then includes experimental errors and the errors due to the predictive model being not the true model. The first kind of errors, the experimental errors, can never be entirely eliminated whatever the precision of the experimental measurements. The modeling error shall also be almost invariably present, since the mathematical models however sophisticated, can rarely describe a complex physical phenomenon exactly. The importance of understanding the nature of these errors is obvious, since model discrimination itself is the statistical processing of the errors. No discrimination problem would be present if errors were not involved.

The errors are conveniently assumed to be normally distributed in most of the discrimination problems. This is not a very far-fetched assumption as far as the experimental errors are concerned.

"Everybody believes in the law of errors, experimenters because they think it is a mathematical theorem, the mathematicians, because they think it is an experimental fact."

However, there does not seem to be any justification for treating the modeling errors to be normally distributed. In fact, it might be more apt to describe the errors due to model form by simpler functions like polynomials. This is the basis of model improvement

in generating spline functions.

Reconsidering the question of modeling errors, it might be argued that if modeling errors were small as compared to the experimental errors, there probably would not be much harm in assigning a normal distribution to the overall errors. The modeling errors would be negligible in case of a model describing the mechanism reasonably well and would be relatively small for all the competitive models. Any model form that has large inherent modeling errors associated with it, would be reduced to a non-competitive position outright and thus would not affect further discrimination between the crucial models.

Besides the advantages of treating the errors to be normally distributed are great because of the analytical tractability of the normal form. Almost the entire present discrimination literature has been developed starting with this assumption. Besides, if the normality assumption is not used, then there arises the difficulty of specifying an alternative distribution. The computational disadvantages far outweigh any advantages that might accrue in terms of better discrimination. It is clear from the foregoing discussion that it is best to live with the assumption of the normality of errors.

However, it might be worthwhile examining the actual errors systematically over the entire operating region for any correlation between the errors and the independent variables. In general, correlated errors signify the existence of large modeling errors. Of course, the errors could also be sometimes correlated due to a poor experimental technique, but these can often be explained by examining the

various measurements and experimental conditions. In such cases, when the model is found to be the cause of correlated errors, the model can be further improved by the addition and or modification of the functions. This can be particularly valuable, where the process is poorly understood or very precise data is available.

It is to be noted that the modeling errors considered above are only errors due to the model form and do not include the errors due to uncertainty in the parameters of a particular model form. The errors in the prediction after $(n-1)$ observations, due to the uncertain parameters are also taken to be normally distributed with mean zero and variance

$$\sigma_{n,i}^2 = \underline{x}_n^t (\underline{X}^t \underline{X})^{-1} \underline{x}_n \sigma^2$$

as described in section 1.3.3. The normality of the errors due to the uncertainty about the parameters, directly follows from the previous assumption of normality about the other two types of errors.

Having decided to treat all errors as normally distributed, we now examine the magnitude of these errors. The mean of these errors is taken to be zero. The variance of the errors may be known or unknown, although, in general it is very hard to have a reasonable estimate of the error variance. While, the experimental errors are amenable to some kind of assessment, the extent of modeling errors is almost never known.

There are standard methods of analysis both when error variance is known or unknown [Raiffa and Schlaifer (1961), Lindley (1965), Quon (1970)]. However, the popular method is to obtain an estimate of the error variance from the least sum of errors squared and treat

it is known in computing the parameter posterior distribution and the expected likelihoods.

Three types of estimates of error variance are commonly employed, viz. (a) A prior estimate obtained from estimating the measurement errors and neglecting the contribution due to the modeling errors. (b) Maximum likelihood estimator of the error variance for a given model, computed as the least sum of errors squared divided by number of data points, that is $\hat{S}^2/n$. This can be interpreted as the average least errors squared per observation and is itself an index of the model performance. (c) The unbiased estimator, computed as the least sum of errors squared divided by the number of observations less the degrees of freedom associated with the model, i.e. $\hat{S}^2/(n-m)$ where m is the number of parameters of the model.

The relative utility of these error variances was examined for a few polynomial model systems. It was found that using the updated unbiased estimate of the error variance resulted in most rapid discrimination followed closely by the results obtained from the maximum likelihoods estimate of the error variance. Using the error variance that was used to generate the data gave a relatively poor discrimination. An illustrative example is given here.

Example 4.1

The following two simple models were used for this example.

$$\text{Model 1.} \quad y^{(1)} = a_1^{(1)} + a_2^{(1)}x$$

$$\text{Model 2.} \quad y^{(2)} = a_1^{(2)} + a_2^{(2)}x + a_3^{(2)}x^2$$

Table 4.1
Results When Discriminating Between the Two Polynomial Models
Using Various Estimates of Error Variance
(Model 2 True)

n	Probability of Model 2 When $\sigma^2 =$		
	1	$\hat{S}^2/n$	$\hat{S}^2/(n-m)$
0	.5	.5	.5
5 (Prior)	.542	.542	.542
6	.589	.632	.644
7	.642	.789	.811
8	.687	.892	.913
9	.808	.948	.963
10	.911	.976	.994

Data was generated from model 2 with $a_1^{(2)} = a_2^{(2)} = a_3^{(2)} = 1$ and error variance of unity. The posterior distribution of the parameters was used to compute the likelihoods. Updated error variance has been used here. Table 4.1 shows the results.

4.1.2 Parameter Joint Distribution

What parameter distribution, prior or the posterior, should be used in computing the expected likelihoods? The question has been debated frequently.

Since the expression for the likelihood contains all of the posterior information, it seems justified to use only the prior information in the parameters to compute the expected likelihood by integrating the likelihood expression with respect to the prior distribution of the parameters. This might even sound analogous to the statement of Bayes' theorem that the posterior probability of a hypothesis is proportional to the product of prior belief about the hypothesis and the posterior likelihood of the data given the hypothesis is true.

On closely examining the problem, it is at once obvious that the Bayes theorem cannot be applied here. The posterior expected likelihood of the data is simply the average of the posterior likelihood function over the entire parameter space. The correctness of the results will depend upon how good is the estimate of the parameter distribution employed in calculating such an average. Clearly, the posterior distribution of the parameters is the best estimate we have to describe the parameter space. Thus, using the posterior

distribution of the parameter sounds more logical and is not equivalent to 'using the information twice'.

Further, the use of prior parameter distribution fails to give any results for an improper uniform prior since the expected likelihoods obtained are zero for all the models. Also, it was suspected that for the vague priors, the model discrimination would be very sensitive to the degree of vagueness if the parameter prior distribution is used. The following two examples were constructed to demonstrate the effect of vague priors.

Example 4.2

Model 1. $y = a_1 + b_1 x$

Model 2. $y = a_2 + b_2 x^2$

Data were generated from $y = 10 + 10 x^{1.2} + \varepsilon$ with known error variance $\sigma_\varepsilon^2 = 1$; 5 points were obtained at $x = 0, 0.5, 1.0, 1.5$ and 2.0 .

Prior parameter distributions for the rival models were

$$\underline{C}_{0,1} = \underline{C}_{0,2} = \begin{bmatrix} 10^4 & 0 \\ 0 & 10^4 \end{bmatrix}; \quad \underline{\mu}_{0,1} = [10, 10];$$

$\underline{\mu}_{0,2}$ was varied to examine the effect on the posterior probabilities of the models. The prior probability of each model was taken to be 0.5.

The posterior probability of model 1 varied from .42 to .96 as $\underline{\mu}_{0,2}$ was given values from $[10, 10]$ to $[200, 200]$ when using the parameter prior distribution. On the other hand, using the parameter posterior distribution, there was practically no effect of different

Table 4.2

Effect of Prior Parameter Values

On the Model Discrimination Between Two Rival Polynomial Models

$\mu_{0,2}$	Posterior Probabilities of Model 1 Using	
	Parameter Prior Distribution	Parameter Posterior Distribution
[10 10]	.420	.261
[20 20]	.428	.261
[50 50]	.467	.261
[100 100]	.630	.261
[200 200]	.965	.262

Table 4.3
Effect of Prior Variance-Covariance Matrix
On the Model Discrimination
Between Two Rival Polynomial Models

$C_{0,1}$		Posterior Probabilities of Model 1 Using	
		Parameter Prior Distribution	Parameter Posterior Distribution
$\begin{bmatrix} 10^4 & 0 \\ 0 & 10^4 \end{bmatrix}$		.106	.365
$\begin{bmatrix} 10^3 & 0 \\ 0 & 10^3 \end{bmatrix}$		.544	.365
$\begin{bmatrix} 10^2 & 0 \\ 0 & 10^2 \end{bmatrix}$		.924	.366

degrees of vagueness on the posterior probabilities. The results are presented in Table 4.2.

Example 4.3

The same competing models were used as in Example 4.2 but data were generated from model 1 with $a_1 = b_1 = 1$ and $\sigma_\varepsilon^2 = 1$, at $x = 0, 0.5, 1.0, 1.5$ and 2.0 . The model parameter priors are

$$\underline{\mu}_{0,1} = \underline{\mu}_{0,2} = [1, 1] \quad \text{and} \quad \underline{C}_{0,2} = \begin{bmatrix} 10^3 & 0 \\ 0 & 10^3 \end{bmatrix}$$

$$\underline{C}_{0,1} \text{ was varied from } \begin{bmatrix} 10^2 & 0 \\ 0 & 10^2 \end{bmatrix} \text{ to } \begin{bmatrix} 10^4 & 0 \\ 0 & 10^4 \end{bmatrix} \text{ and}$$

was found to reduce the model 1 from a posterior probability of 0.924 to 0.106 when using the parameter prior distribution. However, using the parameter posterior distribution gave consistent results as shown in Table 4.3.

Thus, the model discrimination procedure is very sensitive to the parameter prior when using the parameter prior distribution to compute the expected likelihood. In fact, as can be seen from relation 2.13, the expected likelihood of the model is approximately proportional to $|\underline{C}_0|^{-1/2}$ for large $\underline{C}_0$. Hence, if the prior for a particular model is made somewhat more vague, keeping the same priors for the remaining models, the posterior probabilities will be drastically affected. One might obtain almost any set of results by specifying

a somewhat different vague prior. This result lends unequivocal support for using the posterior rather than the prior parameter distribution in computing the expected likelihoods.

Hitherto, the standard Bayesian model discrimination procedure has been to employ parameter prior distribution to describe the parameter space for the likelihood function [Reilly (1970), Box and Hill (1967), Hill (1966), etc]. It might therefore seem surprising that no one has experienced these problems in the face of evident inconsistency. The answer probably is that no one has tried to incorporate these vague priors in the model discrimination problems. The usual practice has been to get around the problem by using part of the data to obtain a prior distribution of the parameters and then continue further by using this prior.

Box and Hill (1967) considered the following four polynomial models for sequential model discrimination.

$$\text{Model 1.} \quad y_1 = a_1 x$$

$$\text{Model 2.} \quad y_2 = a_2 + b_2 x$$

$$\text{Model 3.} \quad y_3 = a_3 + b_3 x + c_3 x^2$$

$$\text{Model 4.} \quad y_4 = a_4 x + b_4 x^2$$

The data was generated from model 3 with known error variance $\sigma^2 = 1$ and $a_2 = b_3 = c_3 = 1$. To start with, all the models were considered equi-likely. Table 4.4 gives the results due to Box and Hill [Hill (1967)] and Table 4.5 gives the results using the same data but employing posterior distribution of parameters.

It is not clear as to what procedure Box and Hill used in arriving at the posterior probabilities for the initial five data points. It is, however, evident that they could not have started with an improper prior because of the problems associated with it while using the parameter prior distribution in computing the expected likelihoods. Examining the posterior probabilities after the initial five data points from Tables 4.4 and 4.5, it appears that Box and Hill used the first five data points to generate a prior and subsequently, they updated the model probabilities by using the prior distribution of the parameters for every additional data point generated. If, this were the case, all the subsequent model discrimination is really being carried among the specific members of the general models obtained after incorporating the first five data points. The use of posterior distribution of parameters also implies discrimination among specific members of the families, but these members represent the "best" member of each family at a given point in the data gathering process. However, the results of model discrimination for the given problem are not significantly different in Table 4.5 by using the posterior distribution of parameters from those reported by Box and Hill in Table 4.4.

Problems associated with the use of parameter prior distribution become still more severe for the non-linear models because of the additional approximations due to linearization. Reilly (1970) suggests the use of posterior parameter means but the prior variance-covariance matrix for obtaining expected likelihoods. This, however, still leaves the problem of hypersensitivity of the results to the

Table 4.4

Results When Discriminating Among the Four Polynomial Models
 Using Prior Parameter Distribution
 In Computing the Likelihood^{*}

n	x	y	$P_{1,n}$	$P_{2,n}$	$P_{3,n}$	$P_{4,n}$
0			.2500	.2500	.2500	.2500
1	0.0	1.431				
2	1.0	2.576				
3	2.0	7.540				
4	3.0	11.765				
5	4.0	20.442	.0024	.0058	.6583	.3335
6	0.0	1.837	.0009	.0012	.8777	.1202
7	0.0	0.140	.0017	.0022	.7471	.2490
8	0.0	1.686	.0007	.0013	.9038	.0942
9	0.0	1.714	.0002	.0009	.9707	.0282

^{*} [Hill (1966)].

Table 4.5

Results When Discriminating Among the Four Polynomial Models
 Using Posterior Parameter Distribution
 In Computing the Likelihood

n	x	y	$P_{1,n}$	$P_{2,n}$	$P_{3,n}$	$P_{4,n}$
0			.2500	.2500	.2500	.2500
1	0.0	1.431				
2	1.0	2.576				
3	2.0	7.540				
4	3.0	11.765				
5	4.0	20.442	.0006	.0006	.6652	.3336
6	0.0	1.837	.0002	.0001	.9099	.0898
7	0.0	0.140	.0003	.0002	.8342	.1653
8	0.0	1.686	.0001	.0001	.9486	.0512
9	0.0	1.714		.0001	.9866	.0133

prior variance-covariance matrix for vague or near vague priors.

One disadvantage of using parameter posterior distribution is the inability to carry out the sequential discrimination by using intermediate parameter distributions as the priors for the next step and update the model posterior probabilities by using few observations at a time. Instead, the prior distribution of the parameters before any data are taken, is used and all the data has to be used together for updating the model probabilities. This could involve somewhat more computations than if parameter prior distribution procedures were used in sequential model discrimination. However, with the more efficient expressions developed here for the expected likelihood in terms of the sufficient statistics of the data, the computations involved for one data point at a time or all the data points together are much the same. In fact, with the improper uniform prior, as often is the case, computations using parameter posterior distribution are almost trivial compared with those required in sequential updating using the prior distribution for the parameters.

4.2 Rationalized Bayesian Model Discrimination Procedure

The recommended model discrimination procedure is outlined below:

- a) Assume the errors to be normally distributed and proceed with the discrimination process. Examine the actual errors in the predictions with the best model. Existence of any correlation of the errors with the independent variables over the entire operating space indicates the ineptness of the model and that it needs to be more sophisticated.

- b) Ordinarily, the discrimination analysis may be carried out assuming the error variance to be known and estimating the error variance from the actual sum of errors squared. For better discrimination, it is desirable to use the updated unbiased estimates of the error variance in computing the parameter posterior distributions and the expected likelihoods. However, for more appropriate results, the error variance should be treated as unknown and its joint distribution with the parameters should be obtained. This joint distribution is then used to compute the expected likelihoods instead of using the joint distribution of only the parameters [see Lindley (1965), Raiffa and Schlaifer (1961)].
- c) It is desirable to use the parameter posterior distribution of parameters to obtain consistent results even when vague or nearly vague priors are used. Furthermore, the use of the posterior distribution of parameters avoids oversensitivity of results to the priors for non-linear models also.
- d) To obtain consistent results while using the posterior distribution of parameters in the sequential discrimination, the sufficient statistics of the entire data must be employed in computing the expected likelihood. The model discrimination should then be carried out with the original priors and this expected likelihood.

5. APPLICATION EXAMPLES

5.1 Petroleum Reservoir Estimation from Material Balance

This chapter is intended to be an exercise and test for the techniques of parameter estimation. Estimation techniques for the parameters of a petroleum reservoir material balance equation (algebraic) and for the parameters in a set of non-linear chemical kinetic differential equations are considered here. We have assumed that the model form is known for certain and that the only uncertainty lies in the values of the parameters.

The nomenclature employed for the material balance equation for a petroleum reservoir is the one recommended by the Society of Petroleum Engineers in 1956 and is detailed by Amyx, Bass and Whiting (1960). The material balance at any time t , for a general case of a petroleum reservoir with an original gas cap and water drive yields

$$\begin{aligned}
 N_p [B_t + B_g (R_p - R_{si}) + W_p - W_i - G_i B_{gi}] \\
 = N [(B_t - B_{ti}) + \frac{B_{ti}}{1-S_w} (C_f + S_w C_w) \Delta P' \\
 + \frac{m B_{ti}}{B_{gi}} (B_g - B_{gi})] + W_e \quad (5.1)
 \end{aligned}$$

where

- B_g - gas formation volume factor at time t
- B_{gi} - gas formation volume factor at $t = 0$
- B_t - total oil formation volume factor
- B_{ti} - initial total oil formation volume factor
- C_f - rock formation compressibility

- C_w - water compressibility
 G_i - cumulative gas injected
 m - ratio of initial gas-cap-gas-reservoir volume to the initial oil in place
 N - initial oil in place
 N_p - cumulative oil produced
 $\Delta P'$ - reservoir pressure decline since $t = 0$
 R_p - producing gas-oil ratio
 R_{si} - initial solution-gas-oil ratio
 S_w - water saturation
 W_e - cumulative water influx
 W_i - cumulative water injected
 W_p - cumulative water produced

The term on the left of equation (5.1) represents net production from the reservoir in barrels at the reservoir conditions and shall be denoted by F . An excellent treatment on the evaluation of the water encroachment term W_e is given by Craft and Hawkins (1959). Van Everdingen and Hurst (1949) developed a procedure to evaluate W_e by superposition of the contributions due to the pressure declines between all the successive time intervals which may be expressed as

$$W_e = C_1 \sum \Delta P \cdot Q(t_D, r_e/r_w)$$

where the summation extends over all the time intervals until time t . ΔP is the average rate of pressure decline at a particular time point in the summation and is usually computed by taking the average of the

pressure declines in the preceding and the following time intervals. Van Everdingen and Hurst (1949) and Craft and Hawkins (1959) have tabulated Q as a function of the ratio of aquifer-reservoir radii r_e/r_w and the dimensionless time $t_D = (\Delta t_D).t'$ where t' is the actual time in intervals starting from the interval under consideration and up to the total time t . Δt_D is a parameter to convert the actual time to the dimensionless units and C_1 is simply a constant.

Equation (5.1) may be simplified by neglecting the rock and water expansion as for saturated reservoirs and becomes

$$F = NE_o + Nm \frac{B_{ti}}{B_{gi}} E_g + C_1 \int \Delta P.Q(t_D, r_e/r_w) \quad (5.2)$$

Here, $E_o = B_t - B_{ti}$ refers to the oil expansion and $E_g = B_g - B_{gi}$ is the gas expansion in bringing them from the reservoir conditions to the reference conditions. F , E_o , E_g , B_{ti} , B_{gi} and ΔP in (5.2) can be obtained from the actual field measurements. $Q(t_D, r_e/r_w)$ may be read from the tables provided Δt_D and r_e/r_w are known. Thus we have five unknowns in (5.2), viz. the initial oil in place N , the ratio of initial gas cap volume to the initial oil rock volume m , C_1 , Δt_D and the ratio of aquifer-reservoir radii r_e/r_w . Semi-empirical methods are available for the estimation of N , m and Δt_D [Amyx, Bass and Whiting (1960)] but are seldom reliable. It is therefore desirable to estimate these parameters from the production history of the reservoir. As can be seen N , Nm and C_1 are the linear parameters and Δt_D and r_e/r_w are the non-linear parameters.

Since the precision of the field data is usually poor, estimation of all the five parameters using the general case of equation (5.2) may lead to rather unrepresentative estimates of the parameters. Besides, it involves more computations particularly in the non-linear search. Hence, any information that simplifies the general equation should be incorporated to reduce the model to fewer parameters. Now, we consider the various reservoir situations in the order of complexity. However, it is implicit that all the following analysis neglects the rock and water expansion.

5.1.1 No Water Drive, No or Known Original Gas Cap

When no original gas cap is present (5.2) yields

$$F = N \cdot E_o \quad (5.3)$$

This, a trivial case, and parameter estimation constitutes obtaining the least squares estimate of N . From a record of n time intervals, we obtain an $n \times 1$ design matrix, such that

$$\underline{X} = \{x_{ij}\} = \{E_o\}_i \quad \begin{matrix} i=1, \dots, n \\ j=1 \end{matrix}$$

Compute

$$\begin{aligned} A &= \underline{X}^T \underline{X} \\ \hat{N} &= \underline{A}^{-1} \underline{X}^T \underline{f} \\ \hat{S}^2 &= (\underline{f} - \hat{N} \underline{e}_o)^T (\underline{f} - \hat{N} \underline{e}_o) \end{aligned}$$

where $\underline{f}$ is the vector containing cumulative production F from the reservoir for the n time intervals and $\underline{e}_o$ is the vector containing E_o . The unbiased variance of the model is estimated by

$$\sigma^2 = \frac{\hat{S}^2}{(n-m)} \quad m=1$$

The variance of the estimate of oil in place $\hat{N}$ is then the only element of the variance-covariance matrix $\underline{C}$ defined by

$$\underline{C} = \sigma^2 \underline{A}^{-1}$$

$$\sigma_{(N)}^2 = C[1, 1]$$

95% confidence interval of $\hat{N} = \pm 2\sigma_{(N)}$.

The case of known gas cap is identical to the above except that E_o is replaced by $E_t = E_o + m \frac{B_{ti}}{B_{gi}} E_g$.

5.1.2 No Water Drive, Unknown m and N

$$F = N(E_o + m \frac{B_{ti}}{B_{gi}} E_g)$$

or

$$\frac{F}{E_o} = N + G \frac{E_g}{E_o} \quad (5.4)$$

Havlena and Odeh (1963) have suggested the following two methods of estimating m and N

- (a) Assume an m and plot $E_o + m \frac{B_{ti}}{B_{gi}} E_g = E_t$ vs F. Adjust m so that a straight line through the origin is obtained.
- (b) Plot F/E_o vs. E_g/E_o so as to obtain a straight line with N as an intercept and G as a slope.

To employ statistical methods of modeling and parameter estimation, the following approach is recommended for the above methods.

(a) Consider the non-linear model

$$F = N(E_o + m \frac{B_{ti}}{B_{gi}} E_g) + C' (E_o + m \frac{B_{ti}}{B_{gi}} E_g)^2 \quad (5.5)$$

where C' is a nuisance parameter and physically represents the curvature of F vs. E_t plot. Non-linear estimation is employed to estimate m using any of the techniques prevalent in literature [Marquardt (1963), Powell (1964)].

However, the criterion for the parameter estimation is to minimize the sum of errors squared subject to keeping the nuisance curvature parameter C' to an acceptably small value. For each value of m , N and C' are estimated by linear least squares. The final estimates of N are obtained by the linear least squares using $C' = 0$ and the estimated value of m . The variance of the fit and the confidence intervals of the linear parameters may be estimated as in section 5.1.1. However, the variance-covariance matrix of the linear parameters shall now include all the uncertainties associated with the non-linear parameter m .

(b) This is a case of a simple linear model.

Let $y = \frac{F}{E_o}$

form $\underline{X}$ such that

$$\begin{aligned} [\underline{X}]_{i,1} &= 1 & i &= 1, n \\ [\underline{X}]_{i,2} &= \begin{bmatrix} E_g \\ E_o \end{bmatrix}_i \end{aligned} \quad (5.6)$$

The parameters and the variance-covariance matrix is obtained as in section 5.1.1. The variance of N and G is given by

$$\sigma_{(N)}^2 = \underline{C}[1, 1]$$

$$\sigma_{(G)}^2 = \underline{C}[2, 2]$$

5.1.3 Water Driven Reservoirs, Two Unknowns

With water drive present but no or known original gas cap, the MBE reduces to

$$\frac{F}{E_t} = N + C' \frac{\sum \Delta P \cdot Q(t_D, r_e/r_w)}{E_t} \quad (5.7)$$

where
$$E_t = E_o + m \frac{B_{ti}}{B_{gi}} E_g$$

There are four parameters in the above model, two linear, N and C and two non-linear, $\frac{r_e}{r_w}$ and Δt_D .

Application of analytical techniques for the estimation of non-linear parameters (r_e/r_w) and Δt_D is limited by computations of $Q(t_D, r_e/r_w)$, the rigorous calculations of which require vast computational efforts at each step. It is therefore desirable to approximate it by an empirical polynomial from the tabulated values. One such computer program 'WDR' that evaluates $Q(t_D, r_e/r_w)$ in the range of $\frac{r_e}{r_w} = 5 - 10$ with an accuracy of $\pm 2\%$ and for $\frac{r_e}{r_w} = 1.5 - 5$, within $\pm 5\%$ is given in the appendix.

The statistical parameter estimation procedure, then, consists of employing a two-dimensional non-linear search for $\frac{r_e}{r_w}$ and Δt_D and simultaneously estimating the linear parameters N and C as indicated by the following steps:

- (a) At a particular iteration, choose or arrive at a value of r_e/r_w and Δt_D .
- (b) Compute the linear parameters N and C for the known values of r_e/r_w and Δt_D by the linear regression described in section 5.1.1.
- (c) Compute the sum of errors squared which now includes the errors due to the linear as well as the non-linear parameters and of course the measurement errors.
- (d) Choose the set of r_e/r_w and Δt_D that leads to minimum sum of errors squared fit for the resulting linear model.

Experience, however, has shown that minimizing the sum of errors squared alone is not adequate. It is also required that the errors be uncorrelated over the time space. To avoid obtaining a misleading fit, it is proposed to use the following five-parameter model

$$\frac{F}{E_t} = N + C_1 \left[\frac{\sum \Delta P.Q(t_D, r_e/r_w)}{E_t} \right] + C_2 \left[\frac{\sum \Delta P.Q(t_D, r_e/r_w)}{E_t} \right]^2 \quad (5.8)$$

The criterion now is to obtain r_e/r_w and Δt_D that minimizes the sum of errors squared subject to keeping C_2 within a small acceptable value ϵ , i.e.

$$\begin{aligned} \min. \hat{S}^2 \\ |c_2| < \epsilon \end{aligned}$$

The variance-covariance matrix for the linear parameters now includes the uncertainties associated with the non-linear parameters.

5.1.4 General Case, Unknown m, N and Water Drive

Havlena and Odeh (1964) reported maximum difficulties in this case owing chiefly to the necessity of obtaining derivatives of $\frac{B_{ti}}{B_{gi}} E_g$ for use in the existing straight line method. This necessitates the availability of extremely accurate data for any reliable results to be obtained.

However, the non-linear estimation of parameters described here does not require any such derivatives and is expected to perform better than the straight line method proposed by Havlena and Odeh (1963). The generalized MBE is

$$\frac{F}{E_o} = N + G \frac{E_g}{E_o} + \frac{C' \sum \Delta P.Q(\Delta t_D)}{E_o} \quad (5.9)$$

This is a five-parameter model with N, G and C' as the linear parameters and r_e/r_w and Δt_D as the non-linear parameters. The parameter estimation and their variance-covariance matrix may be obtained as in Section 5.1.3, except that now the design matrix formed will be of the order nx3 for the linear parameters. Alternatively, a six-parameter model as in Section 5.1.3, may be employed.

$$\frac{F}{E_o} = N + G \frac{E_g}{E_o} + C_1 \left[\frac{\sum \Delta P \cdot Q(\Delta t_D)}{E_o} \right] + C_2 \left[\frac{\sum \Delta P \cdot Q(\Delta t_D)}{E_o} \right]^2 \quad (5.10)$$

The parameters are estimated by minimizing the sum of errors squared subject to keeping C_2 within a small acceptable limit.

5.1.5 Incorporating Prior Information

Frequently, oilmen have some prior knowledge of the oil in place, size of the gas cap and reservoir characteristics from other exploration activities such as volumetric estimates of the oil in place or geologists' personal estimates based on their past experience. Bayesian statistical methods for parameter estimation enable us to combine the prior information with the new production data in arriving at the final results.

If the prior beliefs are only vague in the sense that no particular value for the oil in place seems more likely than the others, infinite variance may be assigned to the various parameters and such a prior knowledge is called the uniform improper prior. Theoretically, such an informationless situation can never be encountered in oil exploration, since anyone, for example, can assert that the oil in place cannot be negative. Nevertheless, the uniform improper prior can be employed if the prior beliefs are rather vague.

The procedure of the parameter estimation and determining the confidence intervals of the parameters for the general case now becomes as follows:

(a) Assign the expected values to the various linear parameters and call it $\underline{b}_o$.

(b) Next, is the more difficult task of quantifying the strength of one's beliefs or estimates. In this case, a 3×3 variance-covariance matrix has to be specified.

$$\underline{C}_0 = \begin{bmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{bmatrix}$$

where c_{11} , c_{22} and c_{33} are respectively the marginal variances of N , G and C_1 . $c_{12} = c_{21}$, $c_{23} = c_{32}$, etc. are the covariances or a measure of correlation between the two parameters. Usually, it is hard to assign any correlation between the various parameters on physical grounds. Hence, it is usual to make all the off-diagonal elements of $\underline{C}_0$ to be zero implying a prior belief of no correlation between the various parameters.

The variances c_{11} , c_{22} , c_{33} may be assigned with relative ease. For example, a geologist may claim that the volumetric estimates of oil in place are 23.2 MMB and he feels about 95% sure that the actual oil in place will not be off from this figure by more than ± 4 MMB. Assuming normal distribution of his beliefs, as we have done for all the measurement errors and the parameters in our statistical analysis, 4 MMB corresponds to two standard deviations and thus prior $\sigma_{(N)} = 2$, giving $c_{11} = \sigma_{(N)}^2 = 4$. Of course making c_{33} infinite or very large values implies that little is known about the parameter C_1 .

(c) Estimate or assign a value to the error variance arising chiefly due to the measurement errors and call it σ_E^2 .

(d) Form the $n \times 3$ design matrix $\underline{X}$ from the production data at n time intervals as described in earlier cases.

Define

$$\underline{A} = \underline{X}^T \underline{X}$$

$$\underline{b} = \underline{A}^{-1} \underline{X}^T \underline{y}$$

$$\underline{A}_0 = \frac{1}{\sigma_E^2} \cdot \underline{C}_0^{-1}$$

where $\underline{y}$ is a vector containing n observed values of the left-hand side of the model, i.e. $\frac{F}{E}_0$.

(e) Estimate the updated values of the parameters and variance-covariance matrix

$$\underline{A}_n = \underline{A}_0 + \underline{A}$$

$$\underline{b}_n = \underline{A}_n^{-1} \{ \underline{A}_0 \underline{b}_0 + \underline{A} \underline{b} \}$$

$$\sigma^2 = \frac{\hat{S}^2}{n-m} \quad m=3$$

A revised estimate of error variance may be obtained by computing a weighted average of σ^2 and σ_E^2 according to certain beliefs in the prior estimate σ_E^2 in terms of equivalent observations. Call this revised estimate $\hat{\sigma}_E^2$. Variance-covariance matrix is now

$$\underline{C}_n = \hat{\sigma}_E^2 \underline{A}_n^{-1}$$

Variance of N , G and C are the respective diagonal elements of $\underline{C}_n$.

This approach can be used to update the estimates of N and G as and when new field data is obtained without resorting to the treatment of entire earlier data which has already been processed. The parameter estimates and the variance-covariance matrix of the previous data is treated as the prior for the new analysis.

5.2 Petroleum Reservoir Estimation--Field Examples

Example 5.1

This is a case of water drive without any gas cap. The example is taken from Van Everdingen et al (1953). An independent volumetric estimate of oil in place is 24-25 MMB.

Since C_2 in equation (5.8) was found to be very small for the case of infinite water drive, the model employed was

$$\frac{F}{E_o} = N + C_1 \left[\frac{\sum \Delta P \cdot Q(t_D, r_e/r_w)}{E_o} \right]$$

It was decided to estimate the parameters by using partial data and compare the results as more and more data was employed. Parameters were estimated by minimizing the sum of errors squared. Since $r_e/r_w = \infty$ was found to give markedly superior fit as compared to some of the other sets of r_e/r_w tried, one-dimensional non-linear search was employed for Δt_D with $r_e/r_w = \infty$. The following polynomial fit was obtained from the tabulated values of $Q(t_D, r_e/r_w)$ for $r_e/r_w = \infty$ [Craft and Hawkins (1959)] and was used here.

$$Q(t_D) = - .251233 + 1.503974t_D^{1/2} + .2955831t_D \\ - .000142643t_D^2 + 1.512 \times 10^{-7}t_D^3 - 1.1496 \times 10^{-10}t_D^4$$

This polynomial was found to give $Q(t_D)$ correct within 1.5% over the range of $t_D = 10$ to 700.

The reservoir started production in 1942 and one data point was available marking the end of each quarter up to the last quarter of 1950. Partial data sets were considered until the end of 1944, 1946, 1948 and the entire data until the end of 1950. Table 5.1 gives the estimates of the initial oil in place N and the variance of these estimates $\sigma_{(N)}^2$ for the various sets of partial data. The estimates of the original oil in place are reasonably good even when only the first three years of data have been used. $\Delta t_D = 14$ was found to give the minimum sum of errors squared using the entire data set the same as reported by Van Everdingen et al, after a complex analysis of the variations in N , calculated at each time interval from the material balance equation. The original oil in place was found to be 25.92 MMB and $C_1 = 0.000193$. Variance-covariance matrix for N and C_1 is

$$\underline{C} = \begin{bmatrix} 6.57 \times 10^{-2} & -1.38 \times 10^{-7} \\ -1.38 \times 10^{-7} & 3.24 \times 10^{-1} \end{bmatrix}$$

This variance-covariance matrix, however, also includes the uncertainties associated with non-linear parameters Δt_D and r_e/r_w .

The variance of fit is $\sigma^2 = 0.23 \text{ MMB}^2$, thus the standard deviation of the estimates of oil in place $\sigma_{(N)} = \sqrt{C[1,1]} = \sqrt{0.0657} = 0.256 \text{ MMB}$ and 95% confidence interval for $N = \pm 2\sigma_{(N)} = \pm 0.513 \text{ MMB}$.

Table 5.2 gives the detailed calculation using the entire data set and $\Delta t_D = 14$. The average pressure decline ΔP , is computed by averaging the pressure decline in the preceding and the following intervals. The dimensionless time $t_D = (\Delta t_D).t'$ is obtained by computing the actual time t' in intervals starting from the given interval and up to the final interval of the data. The polynomial given above was used to get $Q(t_D)$

Example 5.2

We use this example to illustrate the effect of prior information on the final results of Example 5.1. The volumetric estimates of the original oil in place are between 24 and 25 MMB or a mean value of 24.5 MMB. Supposing now that the estimating geologist feels that the actual oil in place lies between 24.5 ± 8 MMB with 95% confidence or 24.5 ± 12 MMB with 99% confidence. Also, he has no idea as to the value of the linear water drive parameter C_1 , to which we arbitrarily assign a value of 1, then

$$\underline{b}_O = [24.5, 1]$$

$$\sigma_{(N)} = \frac{8}{2} = \frac{12}{3} = 4$$

$$\sigma_{(C)} = \infty \approx 1000 \text{ (say)}$$

$$\underline{C}_O = \begin{bmatrix} 16 & 0 \\ 0 & 10^6 \end{bmatrix}$$

Also assume that after examining the measurement techniques for the total cumulative production F , the petroleum engineer assigns

$$\sigma_E = 0.5 \text{ MMB in } F/E_O.$$

Table 5.1
Results for Example 5.1

$r_e/r_w = \infty$

Data Used Up To and Including	Estimates of the Original Oil in Place					
	$t_D = 10$		$t_D = 14$		$t_D = 20$	
	N	$\sigma_{(N)}^2$	N	$\sigma_{(N)}^2$	N	$\sigma_{(N)}^2$
1944	26.8	0.59	26.1	0.31	25.5	0.51
1946	24.4	0.50	25.3	0.27	26.1	0.35
1948	23.9	0.22	25.1	0.13	26.2	0.16
1950	24.7	0.09	25.9	0.07	27.1	0.08

Table 5.2

Details of Calculations for Example 5.1

$$r_e/r_w = \infty, \Delta t_D = 14$$

Year	Quarter	Boundary Pressure @ 8075 ft.	Avg. ΔP for Quarter	$\frac{\Sigma \Delta P Q(t_D) \times 10^6}{T_o}$	E_o	$\frac{F}{E_o} \times 10^{-6}$
	0	3793				
1942	1	3788	2.5			
	2	3774	9.5			
	3	3748	20.0			
	4	3709	32.5			
1943	1	3680	34.0	0.3813	.004225	99.1727
	2	3643	33.0	0.2493	.010032	74.3878
	3	3595	42.5	0.2078	.01759	65.8676
	4	3547	48.0	0.2093	.02432	66.0182
1944	1	3518	38.5	0.2267	.02951	69.6051
	2	3485	31.0	0.2443	.03438	73.2461
	3	3437	40.5	0.2583	.04002	75.5010
	4	3416	34.5	0.2736	.04534	78.7561
1945	1	3379	29.0	0.2786	.05233	78.7521
	2	3358	29.0	0.2989	.05644	83.0141
	3	3338	20.5	0.3241	.05922	88.7914
	4	3329	14.5	0.3473	.06204	92.3293
1946	1	3324	7.0	0.3754	.06347	98.0065
	2	3319	5.0	0.4042	.06455	103.9530
	3	3302	11.0	0.4155	.06835	106.5307
	4	3292	13.5	0.4359	.07056	110.9557
1947	1	3276	13.0	0.4509	.07354	113.8909
	2	3243	24.5	0.4399	.08117	111.5379
	3	3206	35.0	0.4359	.08827	110.2903
	4	3184	29.5	0.4462	.09272	112.3273
1948	1	3165	20.5	0.4517	.09809	113.4767
	2	3135	24.5	0.4506	.10508	112.9093
	3	3108	28.5	0.4546	.11118	113.7178
	4	3095	20.0	0.4706	.11429	116.9875
1949	1	3086	11.0	0.4863	.11721	119.4385
	2	3103	- 4.0	0.5524	.10877	132.8372
	3	3125	-19.5	0.6064	.10380	143.2210
	4	3120	- 8.5	0.6322	.10401	148.1808
1950	1	3115	5.0	0.6582	.10423	152.7882
	2	3114	3.0	0.6842	.10444	157.6309
	3	3115	0.0	0.7113	.10444	162.8010
	4	3116	- 1.0	0.7380	.10444	167.4780

Therefore

$$\underline{A}_O = \sigma_E^2 \underline{C}_O^{-1}$$

$$= \begin{bmatrix} 1.56 \times 10^{-2} & 0 \\ 0 & 2.5 \times 10^{-7} \end{bmatrix}$$

$$\underline{A} = \sigma^2 \underline{C}^{-1}$$

$$= \begin{bmatrix} 3.2 \times 10^1 & 1.36 \times 10^7 \\ 1.36 \times 10^7 & 6.489 \times 10^{12} \end{bmatrix}$$

$$\underline{b} = [25.926 \quad .000193]$$

Now

$$\underline{A}_n = \underline{A} + \underline{A}_O$$

$$\underline{b}_n = \underline{A}_n^{-1} (\underline{A}_O \underline{b}_O + \underline{A} \underline{b})$$

$$= [25.9197 \quad .000193]$$

and

$$\underline{C}_n = \sigma^2 \underline{A}_n^{-1}$$

$$= \begin{bmatrix} 6.5397 \times 10^{-2} & -1.3706 \times 10^{-7} \\ -1.3706 \times 10^{-7} & 3.22653 \times 10^{-13} \end{bmatrix}$$

Suppose that the geologist had somewhat more confidence in his volumetric estimates so that his 95% confidence intervals will correspond to ± 4 MMB, $\sigma_{(N)} = 2$ MMB and now

$$\underline{A}_O = \begin{bmatrix} 6.248 \times 10^{-2} & 0 \\ 0 & 2.5 \times 10^{-7} \end{bmatrix}$$

The first element of $\underline{A}_0$ is fourfold that of the earlier case, $\underline{b}_0$ remains the same. The results are given in Table 5.3 for various confidence intervals of the prior estimates of oil in place.

Example 5.3

Frequently, it is not possible to measure the boundary pressure and sometimes, the exact boundary is not even known. In such cases, the reservoir average pressure is usually employed in the material balance calculations with certain expected reduction in the accuracy. It was decided to examine the effect of using average reservoir pressure in the field case of Example 5.1 and compare the results for the two cases.

The estimation procedure is exactly the same as in Example 5.1 and the optimal results were obtained for the radial drive with $t_D = 31$ and $r_e/r_w = \infty$. The oil in place was now 23.8 MMB against 25.9 MMB obtained in Example 5.1 and the variance of fit was $\sigma^2 = 0.2053$ even lower than the value of 0.23 MMB^2 obtained by using the average boundary pressure. The 95% confidence intervals were also somewhat lower than those obtained in Example 5.1.

From the values of σ^2 and the 95% intervals of the original oil in place, the use of average reservoir pressure did not cause any appreciable decrease in the accuracy of estimation. Since, the volumetric estimates of the oil in place of 24-25 MMB, are somewhere in between, it is not possible to assert the superiority of one or the other set of values.

This indicates that the error in estimates is more due to the errors in the representative pressure measurements than due to the

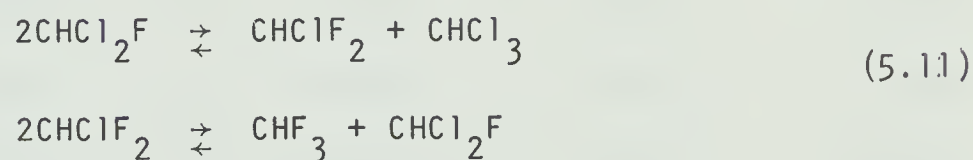
Table 5.3
Effect of Prior Information on the Final Estimates
of the Oil in Place

95% Confidence Intervals for Volumetric Estimates	Posterior Estimates of Oil in Place	Variance of Final Estimates
∞	25.92600	.065689
8.0	25.91965	.065397
4.0	25.90095	.064535
2.0	25.83083	.061305
1.0	25.60882	.051078
0.5	25.16505	.03064
0.0	24.50000	.00000

approximation of using average reservoir pressure for the boundary pressure. In fact, it is felt that weighted average ΔP 's obtained from the boundary pressure and the reservoir pressure will lead to more reliable results than those obtained by using either set of pressures alone.

5.3 Parameter Estimation from a Set of Non-linear Differential Equations

The example chosen represents a contemplated reaction mechanism for the disproportionation of freons on a highly fluorinated alumina catalyst. Stoichiometrically, the reaction may be described by two reversible reactions



Cavaterra, Fattore, and Giordana (1967), studied this reaction at 200°C. Their data is given in Table 5.4. In interpreting the kinetics of this reaction system, Cavaterra et al assumed that the catalytic disproportionation could be represented by the following kinetic model

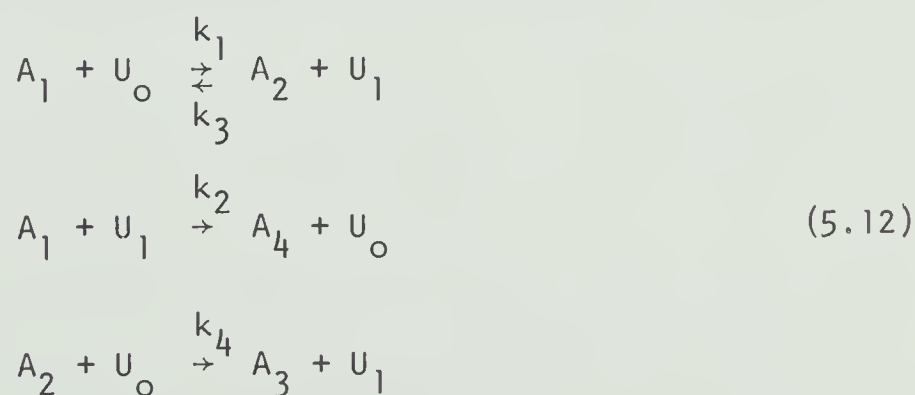


Table 5.4

Disproportionation Reaction of Pure CHCl_2F at 200°C

Time (sec)	Reaction Components (Mol. Fraction)			
	CHCl_2F	CHClF_2	CHF_3	CHCl_3
0.3	0.850	0.063	0.002	0.085
0.5	0.780	0.106	0.004	0.110
1.7	0.435	0.239	0.035	0.291
2.0	0.378	0.255	0.036	0.331
2.3	0.318	0.268	0.047	0.367
2.7	0.289	0.277	0.055	0.379
3.3	0.255	0.286	0.072	0.417
3.7	0.201	0.286	0.078	0.435
4.0	0.176	0.262	0.078	0.484
6.7	0.086	0.215	0.172	0.527
8.0	0.050	0.120	0.235	0.595

where $A_1 = \text{CHCl}_2\text{F}$, $A_2 = \text{CHClF}_2$, $A_3 = \text{CHF}_3$, $A_4 = \text{CHCl}_3$ and U_0 , U_1 are the fluorinating and chlorinating active centers of the catalyst.

The corresponding kinetic model under the steady state conditions may be written as

$$\begin{aligned}\frac{dC_1}{dt} &= -k_1C_1 - k_2C_1 + k_3C_2 \\ \frac{dC_2}{dt} &= k_1C_1 - k_3C_2 - k_4C_2 \\ \frac{dC_3}{dt} &= k_4C_2 \\ \frac{dC_4}{dt} &= k_2C_1\end{aligned}\tag{5.13}$$

Covaterra et al, employed the method of kinetic structural analysis developed by Wei and Prater (1962) to estimate the reaction rate constant. Myint (1970) devised a smoothing scheme and obtained the reaction rates from the smoothed data. Reaction rate constants were then estimated from the resulting algebraic models.

It was decided to employ the quasilinearization procedure to estimate the parameters from the system of simultaneous differential equations. The quasilinearization approach is a second-order iterative process and once within the convergence radius, it converges very fast. However, to obtain a starting value good enough for convergence is a difficult problem. An algorithm proposed by Donnelly and Quon (1970) was employed to arrive within the radius of convergence through the technique of data perturbation. The essential idea is to derive a new

problem so that the existing solution is within the radius of convergence of this new problem and the derived problem is then perturbed towards the original problem in a series of steps, small enough to ensure convergence at each successive step. Donnelly (1968) has written a FORTRAN program for the above algorithm, which was used for the parameter estimation in this example.

With all the four differential equations of (5.13), difficulty was experienced in the final stages of convergence. However, this difficulty disappeared when the last two differential equations of (5.13) were integrated analytically and replaced by the algebraic equations. The results obtained are given in Table 5.5 and compared with the results obtained by Myint (1970) and Cavaterra et al (1967) in Table 5.6. The relative rate constant is obtained by dividing the absolute rate constant by k_1 .

The results obtained by the quasilinearization procedure are reasonably close to those obtained by Cavaterra et al from the more involved Wei and Prater (1962) approach. Myint (1970) also estimated the order of the reaction and therefore solves somewhat different problem. However, it is felt that whatever may be the smoothing function, the smoothed data will not retain the entire information of the raw data. Besides, sometimes it might introduce a bias in the data since the smoothing function may not track the true behaviour of the data generating process.

It might be appropriate to conclude this chapter with a few remarks about petroleum reservoir parameter estimation. It is desirable to consider the most general material balance model and estimate all

Table 5.5
Parameter Estimates of Model (5.13)
Using Data of Table 5.4

I.C. $t = 0$ $C_1 = 1$ $C_2 = 0$ $C_3 = 0$ $C_4 = 0$

Parameter Estimates

$k_1 = 0.2893$ $k_2 = 0.2603$ $k_3 = 0.1521$ $k_4 = 0.1461$

Variance-Covariance Matrix

0.001666	-0.000377	0.001516	0.000463
-0.000377	0.002595	0.002408	-0.003119
0.001516	0.002408	0.006495	-0.004081
0.000463	-0.003119	-0.004081	0.004870

Table 5.6

Comparison of the Parameters Values with

Those Obtained by Myint (1970) and Cavaterra et al (1967)

j	Cavaterra et al (Wei Prater Approach)		Myint		This Work Quasilinearization	
	Relative Rate Constant k'_j	Reaction Order	Relative Rate Constant k'_j	Reaction Order	Relative Rate Constant k'_j	Reaction Order
1	1	1	1	1.168	1	1
2	0.89	1	.8927	1.028	.907	1
3	0.547	1	.3689	1.024	.526	1
4	0.429	1	.2316	0.742	.505	1

the parameters involved. The significance of each term in the general model may then be inferred from the marginal variance of the corresponding parameter. If, however, the computations for the general case are very large, and a very strong a priori information can be used to simplify the model, resort may be made to the simpler cases discussed in this chapter.

6. DISCUSSION

6.1 Summary

The purpose of this thesis has been to develop a procedure for (i) discriminating among the rival models, and (ii) the design of experiments to choose the experimental setting that is expected to result in maximum discrimination among the rival models.

Hitherto the standard Bayesian model discrimination procedure has been to employ parameter prior distribution to describe the parameter space for the likelihood function. The results of model discrimination [Chapter 4] when using parameter prior distribution in computing the expected likelihood was found to be very sensitive to the parameter distribution for vague or near vague priors. Constructed examples were used to demonstrate the relative inconsistency of using parameter prior distribution rather than the parameter posterior distribution.

Also, the expected likelihood of the data is simply the average of the likelihood function over the entire parameter space. The posterior parameter distribution being more accurate representation of the parameter space, should be used to compute the expected likelihoods.

When errors are assumed to be normally and independently distributed with known error variance, the joint distribution of the parameters is multivariate normal. For this case, new expressions [Chapter 2] for the expected likelihood could be developed that involved only the sufficient statistics of the data. The proposed expressions have considerable advantage over the existing method in

terms of the computational savings and reduced round off errors.

Frequently, a situation arises when more than one model can describe a physical phenomenon about equally well. In such instances, it is required to choose an experimental setting that would sharpen the discrimination between the rival models. Model discrimination then becomes a sequential problem of designing the experiments and incorporating the new results for further discrimination. However, when using the posterior parameter distribution to compute the expected likelihood, the entire data must be used against the original parameter prior at each stage of the sequential discrimination.

A discrimination function R was proposed [equation (3.21)], the development of which was based on the concept of entropy as a measure of the uncertainty associated with the system as described by Kulback (1959) and Shannon (1948). The following assumptions were made to make the analysis feasible: (i) the observations are normally and independently distributed with a known and constant error variance; (ii) the model response y_j is linear in parameters $\underline{a}_j$ in the neighbourhood of the least squares estimates of the parameters $\hat{\underline{a}}_j$; (iii) the data is being generated from among the models under consideration.

Several design criteria have been developed in the past few years; notable among them being those proposed by Roth (1965), Box and Hill (1967) and Reilly (1970). However, the criterion proposed in this work evaluates the expected change in entropy or the expected change in the uncertainty of the situation and is both

simple and exact. It also is consistent with the attempts of earlier workers to maximize the difference in the responses of the various models by the inclusion of $(\hat{y}_{n+1,i} - \hat{y}_{n+1})^2$ in equation (3.21) but restrains it with the uncertainty term $(\sigma^2 + \sigma_{n+1}^2)$.

Some illustrative model building examples have been presented in the field of petroleum reservoir parameter estimation in Chapter 5. Parameter estimation from the material balance equation has been discussed for several cases in order of the complexity. Simultaneous non-linear search and linear estimation is recommended.

In addition, an exercise in parameter estimation from a set of kinetic non-linear differential equations has been included in Chapter 5. Quasilinearization procedure together with an algorithm proposed by Donnelly and Quon (1970) was found to give results that are in agreement with those obtained from other methods [Cavaterra et al (1967), Myint (1970)].

6.2 Future Work

One area of future work would be to develop expected likelihood expressions for the case of unknown error variance. The marginal distribution for the error variance is gamma-2. The relative intractability of gamma-2 is the major source of difficulty. However, gamma-2 may be approximated by a normal distribution with the same mean and variance as that of the original distribution itself. The resulting form is expected to be more tractable.

Expected likelihood expressions involving only the sufficient statistics for a multi-response case would be very desirable.

Another area for further development is that of the sequential design of experiments for model discrimination. An expected entropy change criterion for the case of unknown error variance would be very useful. The criterion needs also to be extended to the multi-response situations with the variance-covariance matrix for the various responses both known and unknown. A criterion to select the experimental setting that takes into account both the model discrimination and the precise parameter estimation would be ideal.

There is a great need for a more unified treatment of the various techniques used in model building and model discrimination. The purpose of this treatment would be to indicate the possible strategies and the associated numerical techniques that one might adopt during an investigation. However, more experience is needed in the modeling techniques before such an attempt can be made.

BIBLIOGRAPHY

- Amyx, J.W., Bass, M.B., and Whiting, R.L. (1960). Petroleum Reservoir Engineering, McGraw-Hill, N.Y., Chapter 8, 561-587.
- Box, G.E.P. (1954). The exploration and exploitation of response surfaces, Biometrics, 10, 16.
- Box, G.E.P. (1960). Some general considerations in process optimization, J. Basic Eng., 82, 113.
- Box, G.E.P. (1964). An introduction to response surface methodology, originally prepared for International Encyclopedia of the Social Sciences, Technical Report 33, Department of Statistics, University of Wisconsin, Madison, Wisconsin.
- Box, G.E.P., and Draper, N.R. (1965). The Bayesian estimation of common parameters from several responses, Biometrika, 52, 355.
- Box, G.E.P., and Hill, W.J. (1967). Discrimination among mechanistic models, Technometrics, 9, No. 1, 57.
- Box, G.E.P., and Lucas, H.L. (1959). Design of experiments in non-linear situations, Biometrika, 46, 77.
- Cavaterra, E., Fattora, V., and Giordano, N. (1967). J. of Catalysis, 8, 137.
- Cox, D.R. (1961). Tests of separate families of hypotheses, Proc. Fourth Berkeley Sym., 1, 105.
- Cox, D.R. (1962). Further results on tests of separate families of hypotheses, J. Roy. Stat. Soc., Series B, 24, 406.
- Craft, B.C., and Hawkins, M.F. (1959). Petroleum Reservoir Engineering. Prentice-Hall, N.J., Chapter 5, 204-254.
- Davies, O.L. (1954). The Design and Analysis of Industrial Experiments, Hafner, New York, Chapter 11, 495.
- Draper, N.R., and Hunter, W.G. (1966). Design of experiments for parameter estimation in multiresponse situations, Technical Report 63, Department of Statistics, University of Wisconsin, Madison, Wisconsin.
- Donnelly, J.K. (1968). Identification as a Boundary Value Problem, Department of Chemical Engineering, University of Alberta, Edmonton, Alberta.

- Donnelly, J.K. and Quon, D. (1970). Identification of Parameters in Systems of Ordinary Differential Equations Using Quasilinearization and Data Perturbation, Can. J. Ch.E., 48, 114.
- Graybill, F.A. (1961). An Introduction to Linear Statistical Models, Vol. 1, Chapter 10, p. 197, McGraw-Hill.
- Havlena, D., and Odeh, A.S. (1963). The Material Balance as an Equation of a Straight Line, J. Pet. Tech., 896.
- Havlena, D., and Odeh, A.S. (1964). The Material Balance as an Equation of a Straight Line - Part II, Field Cases, J. Pet. Tech., 815.
- Hill, W.J. (1966). Statistical Techniques in Model Building, Ph.D. Thesis, University of Wisconsin.
- Hunter, J.S. (1958). Determination of optimum conditions by experimental methods, Part II-1, Indus. Qual. Control, 15, No. 6, 16.
- Hunter, J.S. (1959a). Determination of optimum conditions by experimental methods, Part II-2, Indus. Qual. Control, 15, No. 7, 7.
- Hunter, J.S. (1959b). Determination of optimum conditions by experimental methods, Part II-3, Indus. Qual. Control, 15, No. 8, 6.
- Hunter, J.S., and Reiner, A.M. (1965). Designs for discriminating between two rival models, Technometrics, 7, 307.
- Jenkins, G.M., Kisiel, A.J., Price, R.J., and Rippin, D.W., (1963). Preliminary Report, Optimization project, Imperial College.
- Kittrel, J.R., Hunter, W.G., and Watson, C.C. (1966). Obtaining precise parameter estimates for non-linear catalytic rate models, AIChE Journal, 12, No. 1, 5.
- Kulback, S. (1959). Information Theory and Statistics, John Wiley and Sons, Inc., New York.
- Lee, E.S., (1968). Quasilinearization and Invariant Imbedding, Academic Press, New York, p. 84.
- Lindley, D.V. (1965). Probability and Statistics, p. 130, Cambridge University Press, Cambridge, England.
- Marquardt, D.W. (1963). J. Soc. Ind. Appl. Math., 11, 431,
- Myint, A. (1970). Interpretation of Multistep rate Data. M.Sc. thesis, Dept. of Chem. Eng. University of Alberta, Edmonton, Alberta.

- Plackett, R. (1960). Principles of Regression Analysis, Oxford.
- Powell, M.J.D. (1964). An efficient method for finding the minimum of several variables without calculating derivatives, Comp. J., 155.
- Quon, D. (1970). Statistical Decision Analysis, Lecture notes, Dept. of Chemical and Petroleum Engineering, University of Alberta, Edmonton, Alberta.
- Raiffa, H., and Schlaifer, R. (1961). Applied Statistical Decision Theory, Harvard University, Boston, Chapter 13, p. 334.
- Reilly, P.M., (1964). Statistical methods in model discrimination, paper presented at 3rd Symposium on Catalysis, Can. Soc. for Chem. Eng., Edmonton, Alberta
- Roth, P.M. (1965). Design of experiments for discrimination among rival models, Ph.D. thesis, Princeton University, Princeton, N.J.
- Scheffe (1959). The Analysis of Variance, Wiley.
- Shannon, C.E. (1948). A mathematical theory of communication, Bell System Tech. Journal, 27, 379 and 623.
- Smallwood, R.D. (1968). A decision analysis of model selection, I.E.E.E. Trans. Systems Science and Cybernetics, Vol. SSC-4 No. 3, 333.
- Van Everdingen, A.F. and Hurst, W. (1949). The Application of the Laplace Transformation to Flow Problems in Reservoirs, Trans. AIME, 186, 305.
- Wei, J., Prater, C.D. (1962). Advances in Catalysis, 13, 213.
- Williams, E.S. (1959). Regression Analysis, John Wiley and Sons, Inc. New York, N.Y.

APPENDIX

The appendix contains the programs used in this thesis. The notation used in the following programs is described here.

A	Dispersion matrix $\underline{X}^t \underline{X}$
A0	Prior estimate of the dispersion matrix
B	Least squares estimate of the parameters (vector)
BL	Linear parameters (vector) in petroleum reservoir estimation
C	Variance-covariance matrix
D	Determinant, normalizing matrix
DSNM	A function subprogram to set up the design matrix
ER	Errors vector
FET	$(F \div ET)$ of section 5.1
FUCN	A function subprogram to compute the model predictions
INX	Mesh size in the design of experiments
IT	Reference index for the competing models
L,L1,L2	Expected likelihoods
LAM	A parameter in Marquardt's algorithm
M	Number of parameters in the given model
MU0	Prior estimate of the parameters (vector)
N	Number of data points
NM	Total number of competing models
PA,PW	Independent variables
QT	$Q(t_D, r_e/r_w)$ of section 5.1

R	Negative of the expected entropy change
RD	Aquifer-reservoir radii ratio r_e/r_w
SS	Minimum sum of errors squared
SETUP	A function subprogram to specify the prior parameter distribution for the given model
T	Time in intervals
TD	Dimensionless time
VAR	Error variance
X,XX	Design matrices
XP	Normal random number generated from 'SIMULATION'
Y	Vector of observations
YH	Expected value of the prediction from the entire metamodel

A PROGRAM 'INV'

A THIS PROGRAM COMPUTES THE INVERSE OF A NON-
A SINGULAR MATRIX BY GAUSS-JORDON PIVOTING.

```

      VINV[[]]V
    V Z←INV M;I;J;K;P
[ 1]   →3×1(2=ρρM)Λ= / ρiI
[ 2]   →0,ρ[]←'NO INVERSE FOUND'
[ 3]   M←Q(1 0 +ρM)ρ(,Q M),J←1<P←1I←1+ρM
[ 4]   M[K,1;]←M[1,K+K1[ /K+|M[1I;1];]
[ 5]   P←1φP,0ρP[K,1]←P[1,K]
[ 6]   M[;1+ρJ]←~J
[ 7]   →2×1E-30>|M[1;1]÷[ /|,M
[ 8]   M←1φ1φ[1] M-(J×M[;1])◦.×M[1;]←M[1;]÷M[1;1]
[ 9]   →4×10≠J←I-1
[10]   Z←M[;ΔP]
    V

```

A PROGRAM 'DET'

A 'DET' OBTAINS THE DETERMINANT OF A
A SQUARE MATRIX BY CHIO'S METHOD.

```

      VDET[[]]V
    V D←DET A;N;A
[ 1]   H←ρA[1;]
[ 2]   I←0
[ 3]   D←1
[ 4]   →LAB×1N=1
[ 5]   LAL:I←I+1
[ 6]   D←D×A[I;I]
[ 7]   A[I;]←A[I;]÷A[I;I]
[ 8]   L←I
[ 9]   LAD:L←L+1
[10]   A[;L]←A[;L]-A[I;L]×A[;I]
[11]   →LAD×1L<N
[12]   →LAL×1I<N-1
[13]   LAB:D←D×A[N;N]
    V

```


A PROGRAM 'SIMULATION'

A 'SIMULATION' GENERATES NORMAL RANDOM NUMBERS

A REFERENCE: 26.2.23, PG 933, HANDBOOK OF MATH-

A MATICAL FUNCTIONS, U.S. DEPARTMENT OF COMMERCE.

```

      VSIMULATION[ ]V
V SIMULATION;C0;C1;C2;D1;D2;D3;T;P
[1]  C0←2.515517
[2]  C1←0.802853
[3]  C2←0.010328
[4]  D1←1.432788
[5]  D2←0.189269
[6]  D3←0.001308
[7]  P←((1?10000))÷10000
[8]  PP←P
[9]  →MAT× 1P≤0.5
[10] P←1-P
[11] MAT:T←(★(÷(P*2)))*0.5
[12] XP←T-(((C0+C1×T)+C2×T*2)÷(((1+D1×T)+D2×T*2)+D3×T*3))
[13] →TAT× 1PP≤0.5
[14] XP←-XP
[15] TAT:→10
V

```

A PROGRAM 'HST'

A THIS PROGRAM COMPUTES THE EXPECTED LIKELI-

A HOODS FROM THE SUFFICIENT STATISTICS OF THE

A DATA WHEN IMPROPER UNIFORM PRIORS ARE USED.

```

      VHST[ ]V
V HST;M;N
[1]  M←ρB
[2]  N←ρX[ ;1]
[3]  L←((2×VAR)★(-N÷2))×(0.5★(M÷2))×(★(-SS÷(2×VAR)))
V

```


A PROGRAM 'LINE'

A THIS PROGRAM COMPUTES THE LEAST SQUARE ESTIMATES
A AND THE VARIANCE-COVARIANCE MATRIX FOR A LINEAR
A MODEL WHEN ERROR VARIANCE IS KNOWN.

```

      VLIHM[[]]V
V  LIHM
[1]  A←(QX)+.×X
[2]  B←(INV A)+.×(QX)+.×Y
[3]  A←A÷VAR
[4]  C←INV A
[5]  SS←+/(Y-X+.×B)*2
V

```

on PROGRAM 'HSS'

A 'HSS' COMPUTES THE EXPECTED LIKELIHOOD OF A GIVEN
A MODEL FROM THE SUFFICIENT STATISTICS OF THE DATA
A AS DEVELOPED IN CHAPTER 2 OF THIS THESIS. L1 USES
A POSTERIOR DISTRIBUTION AND L2 USES THE PRIOR
A DISTRIBUTION OF THE PARAMETERS.

```

      VHSS[[]]V
V  HSS;AE;AA;AB;AC;L11;AD;DD;N
[1]  N←pX[;1]
[2]  DD←DET(A0+A)
[3]  L11←((2×OVAR)*(−H÷2))*(*(−SS÷(2×VAR)))
[4]  AA←(A0+.×MU0)+A+.×B
[5]  AB←(A0+.×MU0)+2×A+.×B
[6]  AD←((QB)+.×A+.×B)+(QAA)+.×(INV A0+A)+.×AA
[7]  L1←(*−0.5×(AD−(QAB)+.×(INV A0+2×A)+.×AB))×L11
[8]  L1←(L1:(DET(A0+A×2))*0.5)×DD*0.5
[9]  AE←((QB)+.×A+.×B)−(QAA)+.×(INV A0+A)+.×AA
[10] AC←(AE+((QMU0)+.×A0+.×MU0))
[11] L2←L11×((DET A0)*0.5)*(−0.5×AC)÷(DD*0.5)
V

```


A PROGRAM 'MGT'

A THIS PROGRAM IS BASED ON MARQUARDT'S (1963)

A ALGORITHM FOR PARAMETER ESTIMATION FROM A

A NON-LINEAR MODEL.

```

      VMGT[ ]V
V  MGT
[ 1 ]  LAM←0.01
[ 2 ]  IT←0
[ 3 ]  I←( 1 )° . = 1
[ 4 ]  LBB:→10
[ 5 ]  PHI←FUCH LM
[ 6 ]  X←DSHM LM
[ 7 ]  G←(QX)+.×(Y-PHI)
[ 8 ]  SS←(Q(Y-PHI))+.×(Y-PHI)
[ 9 ]  A←((QX)+.×X)+A0
[10 ]  AA←A
[11 ]  D←I
[12 ]  IL←0
[13 ]  LAS:IL←IL+1
[14 ]  D[IL;IL]←A[IL;IL]*-0.5
[15 ]  →LAS×√IL<M
[16 ]  A←D+.×A+.×D
[17 ]  D1←(D+.×(INV(A+LAM×I))+.×D)+.×G
[18 ]  D2←(D+.×(INV(A+(LAM:NU)×I))+.×D)+.×G
[19 ]  B1←B+D1
[20 ]  B2←B+D2
[21 ]  BB←B
[22 ]  B←B1
[23 ]  PHI1←FUCH LM
[24 ]  B←B2
[25 ]  PHI2←FUCH LM
[26 ]  IT←IT+1
[27 ]  S1←(Q(Y-PHI1))+.×(Y-PHI1)
[28 ]  S2←(Q(Y-PHI2))+.×(Y-PHI2)
[29 ]  →THRE×1((1/(BB-B1))<0.0001)
[30 ]  B←B1
[31 ]  B
[32 ]  →TWO×√S1>SS
[33 ]  →LBB
[34 ]  ONE:LAM←LAM÷NU
[35 ]  →LBB
[36 ]  TWO:LAM←LAM×NU
[37 ]  →LBB
[38 ]  THRE:→FOUR×√SS<S1
[39 ]  X←DSHM LM
[40 ]  VAR←S1÷N
[41 ]  CONTINUED ON NEXT PAGE

```



```

[41] A←((QX)+.×X)+A0
[42] C←VAR×(INV A)
[43] →FIVE
[44] FOUR:B←BB
[45] VAR←SS÷N
[46] C←VAR×(INV AA)
[47] FIVE:'PARAMETERS : ';B
[48] 'VARIANCE : ';VAR
[49] 'VAR-COVAR: ';C

```

A PROGRAM 'EXPD1'

A 'EXPD1' COMPUTES THE EXPECTED CHANGE IN

A ENTROPY FOR A HYPOTHETICAL EXPERIMENT.

```

      ∇EXPD1[ ]∇
∇ EXPD1;M;N;IT;I;NM
[1]  N←1
[2]  VARV←YHV←2p0
[3]  I←0
[4]  NM←2
[5]  INX←2
[6]  VAR←10*-3
[7]  PA←PW←2
[8]  LAB:I←I+1
[9]  →LAT×\PW>40
[10] →LAS×\PA>50
[11] IT←0
[12] LAC:IT←IT+1
[13] A0←SETUP IT
[14] YHV[IT]←B FUCN IT
[15] PHI←YHV[IT]
[16] X←B DSNM IT
[17] VARV[IT]←X+.×(INV(A0))+.×(QX)×VAR
[18] →LAC×\IT<NM
[19] YH←+/P0×YHV
[20] VARR←(+/(VARV×P0))+VAR
[21] VTT←(⊗(VARR÷(VARV+VAR)))
[22] R←0.5×(((+/P0×(VTT+(((YHV-YH)*2)÷VARR))))))
[23] R;' PA,PW = ';PA;' ' ;PW
[24] PA←PA+INX
[25] →LAB
[26] LAS:PW←PW+INX
[27] PA←0.2
[28] →LAB
[29] LAT:→\0

```

∇

A APL PROGRAMS FOR PETROLEUM RESERVOIR ESTIMATION.

A PROGRAM 'WDR'

A 'WDR' ESTIMATES THE WATER ENCROACHMENT TERM. THIS

A PROGRAM IS THE APL VERSION OF THE FORTRAN PROGRAM

A OBTAINED FROM DR. BENSTEN.

```

      VWDR[ ]V
V QT←RD WDR TD;BEB;RR;AEA;F2;AA;AB;F1;FX1;FX2;LRD
[1] →LAB10×1RD<99
[2] QT←(0.251233)+(1.503974×TD*0.5)+(0.2955831×TD)
[3] QT←QT+(0.000142643×TD*2)+((1.9512E-7)×TD*3)
[4] QT←QT+((-1.1496E-10)×TD*4)
[5] →LAB2
[6] LAB10:→10
[7] LRD←RD
[8] FX1←(((0.21060396)-0.018830396×LRD)×LRD)
[9] FX2←(((0.19175244)-0.018208508×LRD)×LRD)
[10] FX1←1.42555697+((-1.9619319)+FX1)×LRD)
[11] FX2←2.4981863+((-1.7329829)+FX2)×LRD)
[12] F1←(RD×(0.24778632+(8.1497208E-6)×RD))-0.091510171
[13] F2←(RD×((-1.4646512E-5)+(9.7201383E-8)×RD))+0.0018725209
[14] F1←(F1×RD)-0.15353776
[15] F2←(((F2×RD)+0.042167467)×RD)-0.04562844
[16] RR←((RD*2)-1)÷2
[17] AA←(*FX1)*2
[18] AB←(*FX2)*2
[19] →LAB1×1TD>0
[20] QT←0
[21] →LAB2
[22] LAB1:AA*TD←AA×TD
[23] →LAB3×1AA*TD>112
[24] AEA←F1÷(*AA*TD)
[25] →LAB4
[26] LAB3:AEA←0
[27] LAB4:AB*TD←AB×TD
[28] →LAB5×1AB*TD>112
[29] BEB←F2÷(*AB*TD)
[30] →LAB6
[31] LAB5:BEB←0
[32] LAB6:QT←RR-2×(AEA+BEB)
[33] LAB2:→10

```

V


```

A          PROGRAM      'PRM'

A          'PRM' ESTIMATES BOTH THE LINEAR AND NON-
A          LINEAR PARAMETERS FROM THE MATERIAL
A          BALANCE MODEL AND THE PRODUCTION HISTORY.

```

```

      VPRM[ ] ] V
      V PRM
[ 1 ]      N ← ρFET
[ 2 ]      X ← ( 1N ) ° . = 13
[ 3 ]      X[ ; 1 ] ← N ρ1
[ 4 ]      SEARCH
[ 5 ]      XX ← ( 1N ) ° . = 12
[ 6 ]      XX[ ; 1 ] ← X[ ; 1 ]
[ 7 ]      XX[ ; 2 ] ← X[ ; 2 ]
[ 8 ]      C ← INV( ( λ XX ) + . × XX )
[ 9 ]      BL ← C + . × ( λ XX ) + . × FET
[10 ]      YC ← XX + . × BL
[11 ]      S ← + / ( FET - YC ) * 2
[12 ]      'LINEAR PARAMETER' ; BL
[13 ]      'NONLINEAR PARAMETERS' ; B
[14 ]      VAR ← S ÷ ( N - ( 2 + ρ XX [ 1 ; ] ) )
[15 ]      C ← C × VAR
[16 ]      'LINEAR VAR-COVAR'
[17 ]      C
[18 ]      'ERROR VAR:' ; VAR

```

∇

```

A          PROGRAM      'SEARCH'

A          'SEARCH' PERFORMS A NON-LINEAR SEARCH TO
A          ESTIMATE THE NON-LINEAR PARAMETERS.

```

```

      VSEARCH[ ] ] V
      V SEARCH
[ 1 ]      LAC: S ← + / ( FET - FOFX B ) * 2
[ 2 ]      S ; 'AT ' ; B ; ' AND BL: ' ; BL
[ 3 ]      BB ← B
[ 4 ]      B ← ]
[ 5 ]      → LAT × 1 B [ 1 ] = 0
[ 6 ]      → LAC
[ 7 ]      LAT: B ← BB

```

∇

A PROGRAM 'FOFX'

A THIS PROGRAM COMPUTES ALL THE TERMS IN THE
 A MATERIAL BALANCE EQUATION TO OBTAIN THE
 A DESIGN MATRIX.

```

      VFOFX[ ]V
V  YC←FOFX B;NN
[1]  TR←T×B[2]
[2]  RD←B[1]
[3]  NN←ρT
[4]  QT←NNρ0
[5]  IL←0
[6]  LAA:IL←IL+1
[7]  QT[IL]←RD WDR TR[IL]
[8]  →LAA×\IL<NN
[9]  J←0
[10] LAT:J←J+1
[11] TERM←0
[12] I←0
[13] LAK:I←I+1
[14] TERM←TERM+DP[J+5-I]×QT[I]
[15] →LAK×\I<J+4
[16] X[J;2]←TERM
[17] →LAT×\J<N
[18] X[;2]←X[;2]÷ET
[19] X[;3]←X[;2]*2
[20] AX←(λX)+.×X
[21] DA←(λ3)0.=λ3
[22] I←0
[23] SIN:I←I+1
[24] DA[I;I]←AX[I;I]*-0.5
[25] →SIN×\I<3
[26] AX←DA+.×AX+.×DA
[27] C←DA+.×(INV AX)+.×DA
[28] BL←C+.×(λX)+.×FET
[29] YC←X+.×BL

```

V

B29977